

TECHNICKÁ UNIVERZITA V LIBERCI

Fakulta mechatroniky, informatiky a mezioborových studií

Studijní program: N2612 – Elektrotechnika a informatika

Studijní obor: 3906T001 – Mechatronika

Automatický převod polyfonní melodie do notového zápisu

Automatic transcription of polyphonic melody to notation

Diplomová práce

Autor: **Bc. Marek Boháč**

Vedoucí práce: Prof. Ing. Jan Nouza, Csc.

TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta mechatroniky, informatiky a mezioborových studií
Ústav informačních technologií a elektroniky
Akademický rok: 2009/2010

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Marek BOHÁČ**

Studijní program: **N2612 Elektrotechnika a informatika**

Studijní obor: **Mechatronika**

Název tématu: **Automatický převod polyfonní melodie do notového zápisu**

Z á s a d y p r o v y p r a c o v á n í :

1. Seznamte se s problematikou akustické teorie hudby a s moderními postupy analýzy signálů hudebních nástrojů.
2. Navrhněte program, který analyzuje hudební nahrávku obsahující vícehlasou melodii (hranou jedním nástrojem) a na základě identifikace a rozpoznávání jednotlivých tónů ji převádí do notového záznamu.
3. Přizpůsobte navržený program pro analýzu melodie hrané syntetizátorem a např. malým dechovým souborem.
4. Vytvořte několik (cca 10) nahrávek různých melodií a proveďte na nich vyhodnocení funkčnosti programu.

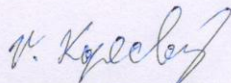
Rozsah grafických prací: Dle potřeby dokumentace
Rozsah pracovní zprávy: cca 40 - 50 stran
Forma zpracování diplomové práce: tištěná/elektronická

Seznam odborné literatury:

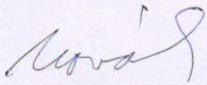
- [1] Grace R.: Hudba a zvuk na počítači. Grada Praha 1999.
- [2] Forró D.: Svět MIDI. Grada Praha 1997.
- [3] Klapuri A.: Signal Processing Methods for the Automatic Transcription of Music. PhD Thesis. Tampere University of Technology. Tampere 2004

Vedoucí diplomové práce: **prof. Ing. Jan Nouza, CSc.**
Ústav informačních technologií a elektroniky

Datum zadání diplomové práce: **16. října 2009**
Termín odevzdání diplomové práce: **21. května 2010**


prof. Ing. Václav Kopecký, CSc.
děkan




prof. Ing. Ondřej Novák, CSc.
vedoucí ústavu

V Liberci dne 16. října 2009

Prohlášení

Byl jsem seznámen s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 - školní dílo.

Beru na vědomí, že Technická univerzita v Liberci (TUL) nezasahuje do mých autorských práv užitím mé diplomové práce pro vnitřní potřebu TUL.

Užiji-li diplomovou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědom povinnosti informovat o této skutečnosti TUL; v tomto případě má TUL právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Diplomovou práci jsem vypracoval samostatně s použitím uvedené literatury a na základě konzultací s vedoucím diplomové práce.

Datum

Podpis

Poděkování

Rád bych poděkoval svým rodičům, bez jejichž podpory během studií i v osobním životě bych tuto práci nemohl napsat. Dále bych chtěl poděkovat vedoucímu práce Prof. Ing. Janu Nouzovi, Csc. za odborné vedení a cenné rady. Zvláštní poděkování patří mé kamarádce Marcelle Vošvrdové za její trpělivost při jazykových korekturách.

Abstrakt

Cílem práce je navrhnout a realizovat program pro převod nahrávky polyfonní hudební melodie do notového zápisu. Poznatky z akustické teorie hudby jsou využity k vytvoření speciálních algoritmů určených k analýze číslicových signálů, které představují zkoumanou nahrávku. Informace z oblasti psychoakustiky byly také využity k definování struktury modelů nástrojů. Program musí nejprve identifikovat časový průběh nahrávky. Je potřeba určit jednotlivé hrané tóny, jejich délku a identifikovat nástroje, kterými jsou tóny interpretovány. Klíčovým problémem této části programu je detekce přítomnosti hrané oktávy. Druhým krokem je převod časového průběhu do notového záznamu, ke kterému jsou zapotřebí informace o taktu a tempu skladby. Algoritmy jsou navrženy se zvláštním zaměřením na analýzu dechových a žesťových nástrojů. Pro tuto skupinu nástrojů je provedeno vyhodnocení úspěšnosti programu. K analýze dalších skupin nástrojů (např. strunných) jsou navrženy dílčí úpravy a rozšíření použitých algoritmů a modelů nástrojů.

Klíčová slova: analýza signálů, vícehlasá melodie, přepis

Abstract

Goal of the work is to design and create a program able of transcription of polyphonic musical melody to notation. Knowledge of musical acoustic theory is used to create special algorithms aimed to analyze digital signals, which represents the analyzed record. Information from field of psychoacoustics was used to define structure of musical instrument models. First of all the program has to identify notation in time-domain. It means it has to mark single played tones (especially their length and pitch) and has to identify musical instrument that was used to interpret it. The main problem of this part of the program is to detect octave presence. The second step of transcription of music is converting time-domain melody knowledge into notation. For this part of solution we need information about the beat and time of the music composition. Algorithms have been designed with special orientation to analysis of wind and brass music instruments. For this group of instruments has been made program reliability analysis. To analyze other groups of music instruments (e.g. stringed) there are designed particular changes and extensions of used algorithm and models of instruments.

Keywords: signal analysis, polyphonic melody, transcription

Obsah

1	ÚVOD A MOTIVACE.....	8
2	STANOVENÍ CÍLŮ PRÁCE	9
3	HUDBA A CENTRÁLNÍ NERVOVÁ SOUSTAVA	10
3.1	Lidské ucho	10
3.2	Mozek – zkušenosti a znalosti.....	12
4	HUDBA A POČÍTAČE	13
4.1	WAVE.....	13
4.2	MIDI.....	14
4.2.1	Struktura standardního souboru MIDI	14
4.2.2	Zprávy jazyka MIDI.....	14
4.2.3	Meta-eventy	14
5	POUŽITÉ ANALYTICKÉ METODY.....	16
5.1	Fourierova transformace	16
5.2	Okénkování	18
5.3	Doplnění signálu nulami	20
5.4	Omezení metod a optimalizace	22
5.5	Korelace	24
5.6	Energie konečného signálu	24
6	VRSTVY PŘEVODU MELODIE	25
6.1	Identifikace melodie v časové oblasti	26
6.1.1	Rozlišení pauz a not	26
6.1.2	Určení fundamentální frekvence – PSICE	29
6.1.3	Analýza potřeby modelů nástrojů.....	33
6.1.4	Modely nástrojů.....	35
6.1.5	Identifikace oktáv a nástrojů	37
6.1.6	Souvislost počtu nástrojů a vzniklých hypotéz	39
6.1.7	Návrhy rozšíření algoritmů	40
6.2	Odhad tempa a taktu.....	42
6.3	Převod melodie do notového zápisu.....	44
7	MANUÁL K PROGRAMU	45
7.1	Načtení souboru	45
7.2	Provedení analýzy	46
7.3	Nastavení parametrů analýzy	47
7.4	Export výsledků	48
7.5	O aplikaci	48
8	VYHODNOCENÍ ÚSPĚŠNOSTI PROGRAMU	49
8.1	Odhalené chyby a návrhy řešení	49
8.2	Porovnání úseku předlohy a automatického přepisu.....	50
	ZÁVĚR	52

Seznam zkratk a symbolů.....	53
Seznam použité literatury.....	54
Seznam příloh	55
Příloha 1: Frekvenční analýza délek not.....	1
Příloha 2: Použitý soubor modelů.....	3
Příloha 3: CD	4
Notové zápisy a nahrávky	4
Program.....	4
Zpráva	4

1 Úvod a motivace

V dnešní době se počítače uplatňují téměř ve všech odvětvích lidské činnosti. Trendem se stává komunikace s počítačem pomocí hlasového vstupu. Rozvíjejí se i další formy počítačového vnímání. Problematika počítačového zpracování řeči je velmi rozvinutou oblastí, u níž lze v brzké době očekávat nasazení těchto technologií do běžného života. Kromě vytvoření komfortnějších uživatelských rozhraní umožní ovládat počítač hendikepovaným lidem. Další oblastí počítačového vnímání, která zaznamenala obrovské pokroky, je oblast počítačového vidění. V dvourozměrných aplikacích existují velmi účinné postupy, které umožňují rozpoznávání útvarů, měření rozměrů nebo „čtení“ textu pomocí algoritmů OCR. Vývoj probíhá v oblasti trojrozměrného vidění.

Méně prozkoumanou oblastí je počítačové zpracování hudebních nahrávek. Snazší úlohou je analýza jednohlasé melodie. Tuto problematiku lze považovat za vyřešenou. Vzhledem k tomu, že jsem se tímto tématem zabýval ve své bakalářské práci, bylo pro mě velmi lákavé pokračovat v oboru i diplomovou prací. Analýza polyfonní hudební nahrávky je záležitostí mnohem zajímavější a obtížnější. Doposud existují pouze osamocené pokusy o vytvoření potřebných algoritmů. Byly podniknuty experimenty s trénováním neuronových sítí, aplikace statistických přístupů i použití „klasické“ signálové analýzy. Tyto snahy vedly k dílčím úspěchům na úrovni práce s jedním nástrojem (v drtivé většině případů počítačově syntetizovaným klavírem).

V rámci diplomové práce jsem se pokusil posunout tyto úvahy o něco dál a to hned ve dvou směrech. Jednak se zabývám analýzou nahrávky, v níž lze očekávat přítomnost několika různých nástrojů – je zapotřebí identifikovat tyto nástroje. Druhou významnou změnou je volba nástrojů, které jsou k experimentům použity. Pouhým porovnáním frekvenčních spekter klavíru a dechových nástrojů lze zjistit značné rozdíly mezi nástroji, které kladou výrazně větší nároky na robustnost použitých algoritmů.

2 Stanovení cílů práce

Cílem této práce je návrh programu, který analyzuje nahrávku, obsahující vícehlasou hudební melodii a na základě identifikace a rozpoznání jednotlivých tónů ji převádí do notového záznamu. Dalším úkolem je úprava programu pro analýzu melodie hrané syntetizátorem či malým dechovým orchestrem.

Prvním krokem je rešerše současného stavu znalostí a známých řešení. Zprvé je zapotřebí seznámit se s obdobně zaměřenými pracemi vytvořenými v zahraničí. Zadruhé jde o oblast psychoakustiky, jejíž znalost je nutná k realizaci diplomové práce. Na základě poznatků z těchto oblastí je možné formulovat vlastní postup řešení úlohy a vyhnout se zbytečnému opakování chyb předchůdců. Obzvláště zajímavá je v tomto ohledu disertační práce A. Klapuriho z Tampere University of Technology [6].

Hlavním těžištěm práce je úloha identifikace hraných tónů. Cílem je vytvořit co nejrobustnější algoritmus, který by kromě běžně analyzovaného klavíru (obvykle se jedná o počítačovou syntézu) spolehlivě detekoval i tóny generované co nejširším spektrem skutečných nástrojů. Druhým neméně důležitým tématem je problematika detekce oktáv a konečně možnosti identifikace nástrojů, které interpretují jednotlivé party. Další bodem vypracování je postup odhadu tempa skladby a taktu, ve kterém je původní melodie zapsána. V příslušných kapitolách bude blíže vysvětleno, že se jedná pouze o odhad. Různé kombinace taktu, tempa a délek not mohou vést na stejný časový průběh melodie. Řešení úlohy je proto nejednoznačné.

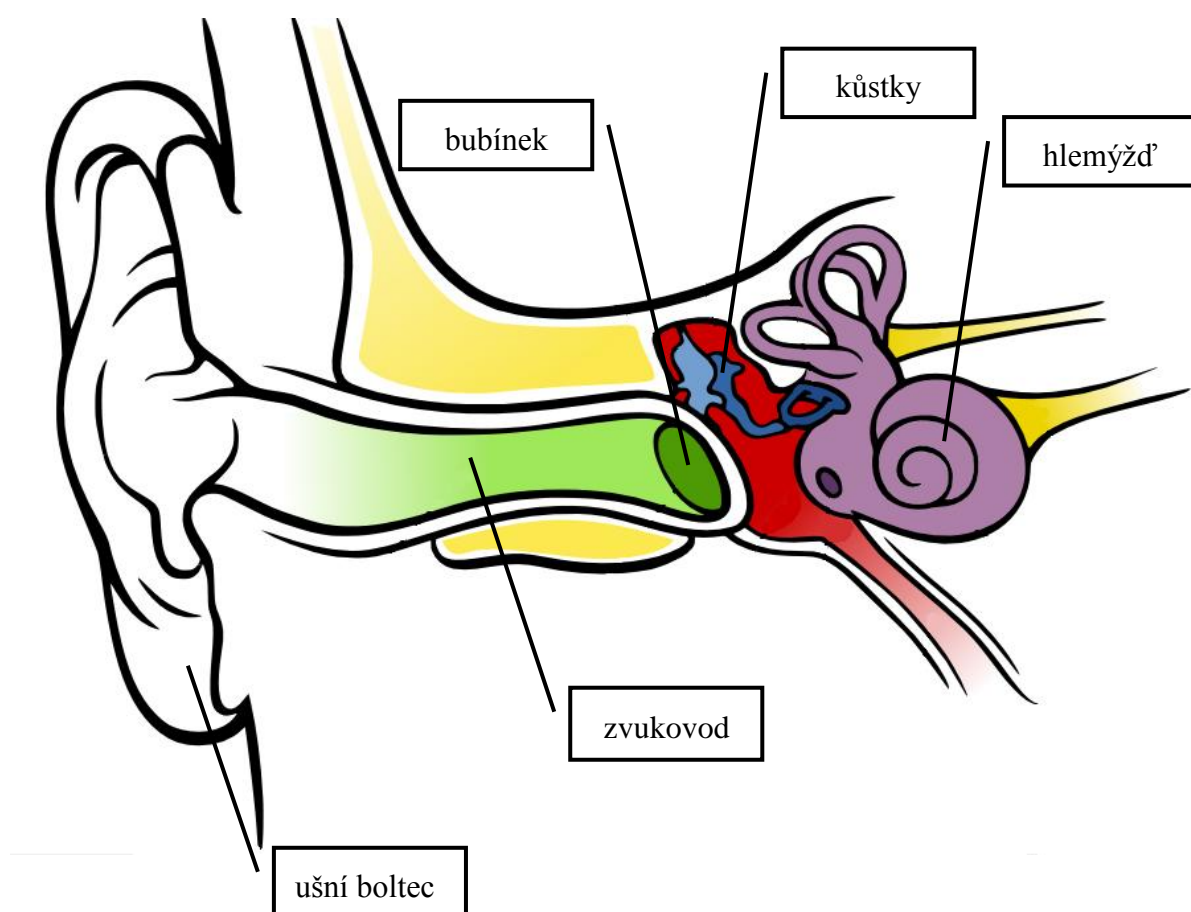
Výstupem programu byl určen standardní soubor MIDI a jednoduchá vizualizace výšek detekovaných tónů v závislosti na čase. Takový formát umožní získat představu o výsledcích analýzy i člověku bez hudebního vzdělání. Formát MIDI umožňuje nejen přehrávání výsledků převodu, ale i jejich zobrazení a úpravu pomocí volně dostupných programů.

Závěrečným krokem práce je vytvoření sady testovacích nahrávek. S jejich pomocí je vyhodnocena úspěšnost programu při převodu melodie do hudebního záznamu. Na základě analýzy úspěšnosti programu a chyb, kterých se dopustil, jsou v závěru navrženy možnosti dalšího rozvoje programu.

3 Hudba a centrální nervová soustava

Na první pohled by se mohlo zdát, že lidský mozek a počítačové zpracování signálů spolu příliš nesouvisí. Opak je ale pravdou. Doposud nejúspěšnějším a nejrobustnějším systémem schopným převádět hudbu do notového zápisu je totiž lidský mozek. Je proto logické věnovat pozornost jeho studiu a tyto poznatky použít při stavbě algoritmů, které by ho měly napodobit. Zkoumání interakcí mezi hudbou a centrální nervovou soustavou lze rozdělit do dvou úrovní. První je otázka funkce lidského ucha, jakožto velice specifického receptoru. Druhým okruhem otázek jsou postupy, které při analýze zvukového vjemu mozek používá.

3.1 Lidské ucho



Obr. 1: Lidské ucho (převzato z [11])

Lidské ucho je komplikovaný systém, který lze poměrně těžko modelovat. Funkci ucha lze stručně popsat následujícím způsobem. Zvuková vlna je zachycena ušním boltcem a postupuje zvukovodem k bubínku, který se jejím působením rozkmitá. Vibrace bubínku se přenášejí přes soustavu kůstek, pružnou blanku a oválné okénko do vnitřního ucha. Vnitřní ucho je vyplněno tekutinou. Jeho hlavní částí je hlemýžď. Mechanické vibrace se šíří tekutinou v hlemýždi a způsobují vychýlení bazilární membrány, která se nachází po celé délce hlemýždě. Tato membrána má specifické mechanické vlastnosti, díky kterým se v každém místě vychyluje v závislosti na frekvenci vlnění. Tím je umožněno rozlišení výšky tónů. Na bazilární membráně jsou uložena ukončení rozvětveného sluchového nervu, kde dochází k přeměně kmitů na elektrické impulsy. Tyto impulsy jsou pak vedeny do mozku.

Nejběžnější postupem, který se užívá k modelování lidského ucha, je řazení sady hřebenových filtrů. Při modelování se využívají poznatky z tzv. psychoakustiky. Psychoakustika je věda, která se zabývá vlivem zvuku na lidský organismus. Model ucha lze odvodit jako nápodobu jednotlivých nervových napojení umístěných na hlemýždi. Každé z vláken nervu je buzeno právě ve chvíli, kdy bazilární membrána pod ním „rezonuje“. Šířky pásem filtrů jsou určovány tak, že člověk dokáže rozeznat přinejmenším intervaly čtvrtiny tónu, což definuje určitý počet filtrů na dekádu. Každému filtru by mělo být přiděleno určité zesílení. Velikost zesílení byla zjištěna v rámci psychoakustických experimentů. Vnímání intenzity zvuku je závislé jednak na frekvenci zvuku a také na hladině hlasitosti akustického signálu.

Pro účely analýzy hudebních nahrávek je mnohem zajímavější skutečnost, že ucho není schopno vnímat změnu fázového rozdílu mezi jednotlivými frekvencemi, ze kterých je zvuk složen. To znamená, že dva zvuky s velmi odlišným časovým průběhem mohou vyvolat stejný akustický vjem. Této skutečnosti lze s výhodou využít například při tvorbě modelů nástrojů. Nemusíme brát v úvahu fázový posun jednotlivých složek spektra, takže stačí použít pouze amplitudy těchto složek. Další užitečnou informací je rozsah frekvencí, které lidské ucho slyší. Víme-li, že ucho rozlišuje zvuky v rozsahu $20 \text{ Hz} \div 20 \text{ kHz}$, lze omezit rozsah pásma, které podrobíme analýze.

3.2 Mozek – zkušenosti a znalosti

Jedním z důvodů, proč hudebník dosahuje při přepisu hudební nahrávky do notového záznamu vyšší úspěšnosti než dosavadní vytvořené programy, je jeho „know-how“. Hudebník během své kariéry získává informace hned v několika oblastech. První oblastí jsou „modely nástrojů“. Každý hudební nástroj má svou specifickou barvu - barva tónu je dána kombinací vyšších harmonických a jejich amplitud. Jednotlivé nástroje se mohou mírně odlišovat. Člověk je díky soustavnému pozorování schopen vytvořit si nejen „širokou databázi modelů“, ale také představu o tom, jaké odchylky můžou v rámci modelu nastat. Druhou výhodou profesionálního hudebníka oproti počítači je znalost harmonie. Jedná se o znalosti o četnosti výskytů určitých akordů – výskyt některých kombinací tónů je mnohem pravděpodobnější než u jiných kombinací. Tyto informace umožňují hudebníkům korigovat chybně identifikované tóny. Poslední skupinou znalostí, které mohou hudebníci využívat, je častý výskyt některých rytmických skupin, které jsou typické pro určité žánry. Navíc může hudebník přepisovat skladbu, kterou již dříve interpretoval buď ve stejné, nebo podobné aranži. V takovém případě je nasnadě, že hudebník je vůči počítači ve výrazné výhodě.

Pokud by se podařilo všechny tyto zdroje informací zapojit do analýzy, mohli bychom oprávněně očekávat zvýšení přesnosti výsledků. Aplikace výše zmíněných principů představuje sama o sobě velmi náročný problém. V rámci mé diplomové práce bude tato část analýzy zjednodušena na úroveň heuristických pravidel, která jsou popsána v dalších pasážích práce.

4 Hudba a počítače

Poslech hudebního díla vyvolává u většiny lidí různě silné emocionální reakce. Chceme-li někomu předat co nejpřesnější informaci o hudebním díle, stačí mu dát partituru skladby, jejíž forma se od dob W. A. Mozarta nijak výrazně nezměnila. Pro počítač je však hudba něčím naprosto odlišným. Nejenže u počítače těžko vyvoláme emoce, ale papír a inkoust nejsou obvyklým médiem pro přenos informace. Pro počítač je hudba pouze dalším balíkem dat. Proto se v této kapitole budu zabývat výběrem vhodných datových formátů a jejich stručnou charakteristikou. Rozbor formátů není těžištěm práce, proto bude rozsah kapitoly co nejvíce zestručněn. V případě zájmu můžete obsáhlejší informaci najít například v mé bakalářské práci [1].

4.1 WAVE

Formát WAVE je podmnožinou standardu RIFF, který je primárně určen k práci s multimediálními soubory. Každý soubor v tomto formátu je složen ze dvou částí. První částí je hlavička souboru. Ta v sobě nese především informace o vzorkovací frekvenci, s jakou byla nahrávka pořízena, o počtu kanálů záznamu a tzv. bitové hloubce. Druhou částí souboru je blok dat, ve kterém jsou přímo uloženy jednotlivé hodnoty navzorkované A/D převodníkem zvukové karty.

Formát WAVE byl zvolen k vytvoření nahrávek hned z několika důvodů. Jednak umožňuje vybrat vzorkovací frekvenci, s jakou jsou pořízeny zkoumané nahrávky. S ohledem na rozsah frekvencí, které dokáže vnímat lidské ucho, jsou vhodné volby 44,1 kHz nebo 48 kHz, popřípadě vyšší. Význam bitové hloubky b lze interpretovat jako počet $N = 2^b$ úrovní, mezi které se lineárně rozdělí množina bodů mezi minimální výchylkou membrány mikrofону (-1) a maximální výchylkou (+1). Jako dostačující se ukazuje $b = 16$ bitů. Podstatnou vlastností tohoto formátu je, že se nepokouší data nijak komprimovat. Díky tomu můžeme předpokládat, že nahrávka bude ovlivněna pouze frekvenčními vlastnostmi mikrofону a A/D převodníku – zvukové karty. V počítači již nedochází k dalším úpravám, které se dějí například u formátu MP3. Pro potřeby této práce je volen počet kanálů 1, tzv. MONO.

4.2 MIDI

MIDI je nejrozšířenější průmyslový standard v oblasti hudební elektroniky. Jeho první specifikace byly zveřejněny v roce 1983. Jedná se o sériový jednosměrný komunikační protokol pracující na vlastním hardwarovém rozhraní. Umožňuje propojování syntetizérů, samplerů, počítačů, ale i scénické techniky v reálném čase. MIDI můžeme chápat na dvou úrovních. První je hardware rozhraní, druhou pak komunikační protokol, který se používá i v rámci jednoho zařízení. Komunikace na platformě MIDI může vznikat buď v reálném čase, nebo může být vytvořena předem a uložena jako standardní soubor MIDI (SMF). Díky své struktuře umožňuje soubor MIDI uložení celé partitury v počítači. Z tohoto důvodu byl vybrán jako výstupní formát programu.

4.2.1 Struktura standardního souboru MIDI

Každý soubor formátu MIDI se skládá ze dvou částí. První je hlavička samotného souboru. Ta obsahuje informace specifikující podtyp formátu, délku hlavičky a vztažnou hodnotu pro kódování tzv. delta-času.

Druhou částí jsou jednotlivé stopy. Každá z nich je uvedena vlastní hlavičkou, jejímž hlavním parametrem je délka příslušné stopy. Stopy jsou složeny z jednotlivých zpráv jazyka MIDI a z meta-eventů.

4.2.2 Zprávy jazyka MIDI

Zprávy jazyka MIDI tvoří základ celého komunikačního protokolu. Skládají se ze stavového bytu a z bytů datových. Stavový byte odlišuje svou hodnotou zprávy od dalších součástí komunikace, rozlišuje data jednotlivých kanálů a systému a identifikuje adresáta zprávy. Význam zprávy je dán obsahem datových bytů. Lze říci, že zpráva je pokynem k provedení nějakého úkonu.

4.2.3 Meta-eventy

Meta eventy jsou kódy pro deklarování zvláštních dat skladby. Jedná se např. o tempo, takt, text, předznamenání, zapnutí a vypnutí jednotlivých not.

Tab. 1: Přehled základních meta-eventů

<i>Název</i>	<i>Kód (hexadecimálně)</i>
Tempo	\$FF \$51 \$03 tt tt tt
Metrum	\$FF \$58 \$04 nn dd cc bb
Předznamení	\$FF \$59 \$02 bk dm
Nota zapnuta	\$90 vt rz
Konec stopy	\$FF \$2F \$00

V hudbě se tempo udává buď slovním údajem, nebo jako počet dob za minutu (beats per minute). V tanečním sportu se tempo definuje počtem taktů za minutu (bars per minute). Ve formátu MIDI je tempo dáno počtem mikrosekund připadajících na jednu čtvrtovou notu.

Metrum, čili takt skladby, je určeno dvěma celými čísly, která tvoří zlomek na začátku notového řádku. Další informací obsaženou v meta-eventu „metrum“ je počet synchronizačních tiků, které uplynou mezi dvěma údery metronomu.

Předznamení je zakódováno ve dvou datových bytech. Počet křížků, respektive béček, je určen číslem v rozsahu $-7 \div 0 \div 7$, kde záporné číslo značí béčka a kladné křížky. Další byte nabývá hodnoty 1 pro mollovou stupnici, zatímco Durová je označena číslem 0.

Meta-event „nota zapnuta“ obsahuje jednak informaci o výšce hraného tónu a také údaj o „rychlosti“ zapnutí. Výška tónu nabývá hodnot $0 \div 127$, což odpovídá teoretickému rozsahu více než deseti oktáv. Rychlost zapnutí původně definuje sílu stisku klávesy klaviatury, což je dáno historickým vývojem normy. Stisk klávesy s nulovou rychlostí se interpretuje jako vypnutí daného tónu. Při zapínání a vypínání not lze s výhodou použít tzv. trvajících stavových byte. Pokud odešleme na sběrnici zprávu „nota zapnuta“, všechny následující datové byty a údaje tzv. delta-času se interpretují, jako kdyby jim předcházel tentýž stavový byte.

Konec stopy je kód, kterým se označuje konec stopy. V případě formátu nula označuje zároveň konec souboru, u dalších formátů odděluje jednotlivé stopy.

5 Použité analytické metody

Kapitola popisuje základní metody analýzy číslicových signálů, na jejichž základě jsou vytvořeny pokročilejší algoritmy uplatněné v programu. Všechny metody byly obsáhle popsány v mnoha publikacích. Teoretické partie budou sepsány co nejstručněji. Nejdůležitější částí kapitoly je shrnutí omezení, která jsou s analýzou číslicových signálů spojena a důsledky těchto omezení na volbu parametrů framu určeného k analýze.

5.1 Fourierova transformace

Průběh signálu se obvykle zobrazuje v závislosti na čase (v časové oblasti). Další možností, jak signál zobrazit, je převést ho do oblasti frekvenční. To znamená, že signál je rozložen na harmonické funkce, jejichž amplitudy a fázová posunutí jsou zobrazena v závislosti na frekvenci. Rozklad obecného signálu (periodického i neperiodického) se realizuje pomocí takzvané Fourierovy transformace. Pro spojitý signál je přímá a inverzní Fourierova transformace dána následujícími vztahy:

$$X(\omega) = F\{x(t)\} = \int_{-\infty}^{+\infty} x(t) \cdot e^{-j\omega t} dt \quad (5.1)$$

$$x(t) = F^{-1}\{X(\omega)\} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(\omega) \cdot e^{j\omega t} d\omega \quad (5.2)$$

$X(\omega)$ je signál transformovaný do frekvenční oblasti, $x(t)$ reprezentuje časový průběh, ω je úhlová frekvence a t je čas. V kontextu této práce budeme ale pracovat s číslicovým signálem. Ten vznikne ze spojitého signálu dvěma procesy – vzorkováním a kvantováním. Vzorkování je výběr hodnot spojitého signálu v ekvidistantně rozmístěných okamžicích, kdy časové okamžiky splňují podmínku $t = k \cdot \Delta t$, kde Δt je perioda vzorkování, k je přirozené číslo. Vzorkovací frekvence je dána jako převrácená hodnota periody vzorkování $f_s = 1/\Delta t$. Kvantování je proces, kdy je obecně libovolný počet úrovní spojitého signálu převáděn do konečné množiny diskrétních hodnot. Je-li signál vzorkovaný a periodický, lze obecnou Fourierovu transformaci přepracovat na diskrétní Fourierovu transformaci (DFT). Pokud je za jednu

periodu signálu naměřeno N hodnot (vzorků), pak perioda signálu (T) a interval vzorkování (Δt) jsou ve vztahu $T = N \cdot \Delta t$. Z předpokladu periodicity vyplývá, že signál periodicky pokračuje před zkoumaným záznamem i za ním. Diskrétní Fourierova transformace (DFT) je pak definována vztahem

$$[F_k]_{BK} = \frac{1}{N} \sum_{i=0}^{N-1} x_i e^{-j \frac{2\pi}{N} ki}, k = 0, 1, 2, \dots, N-1 \quad (5.3)$$

Posloupnost komplexních koeficientů F_k nazýváme Fourierovo spektrum. Vzdálenost sousedních spektrálních čar (Δf) lze určit z následujícího vztahu:

$$\Delta f = f_s / N \quad (5.4)$$

Diskrétní Fourierova transformace podléhá dvěma hlavním omezujícím vlivům. Prvním z nich je předpoklad periodicity signálu. Pokud totiž neprovádíme transformaci úseku signálu, který přesně odpovídá celočíselnému násobku periody, algoritmus si „domyslí“ odlišný periodický průběh signálu. Ve frekvenční oblasti se to projeví jako tzv. rozmazání spektra. O možnostech redukce rozmazání se zmiňuji v kapitole 5.2 *Okénkování*. Druhým omezujícím faktorem je Nyquistův (nebo také Shannon-Kotelnikův) teorém. Ten říká, že nejvyšší frekvence, vyskytující se ve spojitém signálu musí být menší než polovina vzorkovací frekvence. Jev, který nastane při nedodržení tohoto teorému, se nazývá aliasing. Jeho řešení bývá provedeno na úrovni dolnoproustního filtru na vstupu A/D převodníku.

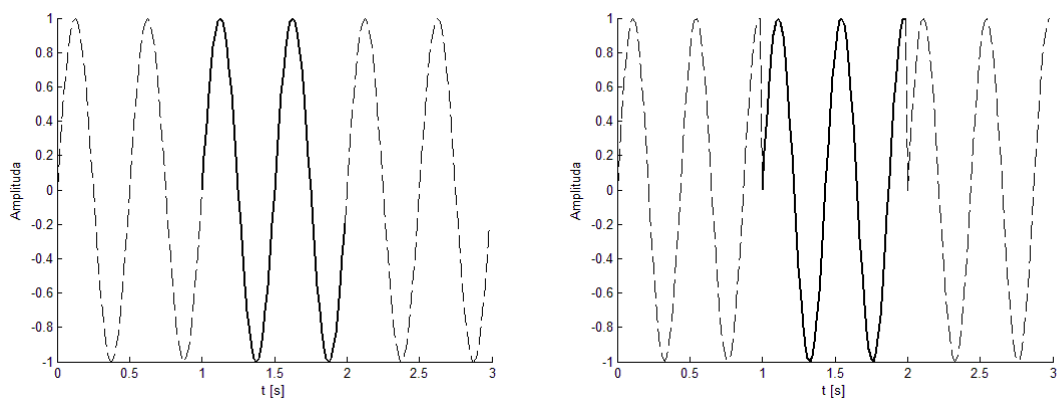
Posledním pojmem, který je třeba zmínit je FFT. Fast Fourier Transform je algoritmus výpočtu DFT, který nevychází z definičního vztahu, ale je optimalizován pro snížení výpočetní náročnosti. Jedním z principů, na kterém může být založen je tzv. decimace v čase, kterou vyvinuli Cooley a Tukey. Algoritmus FFT dosahuje nejlepší úspory výpočtů pro délky analyzovaného vektoru, které jsou mocninou dvojky. Tyto úpravy algoritmu představují revoluční změnu v signálové analýze. Umožňují totiž provádět výpočet koeficientů DFT až o dva řády rychleji (místo minuty trvá výpočet necelou sekundou). Míra „zrychlení“ výpočtu závisí na délce vektoru.

5.2 Okénkování

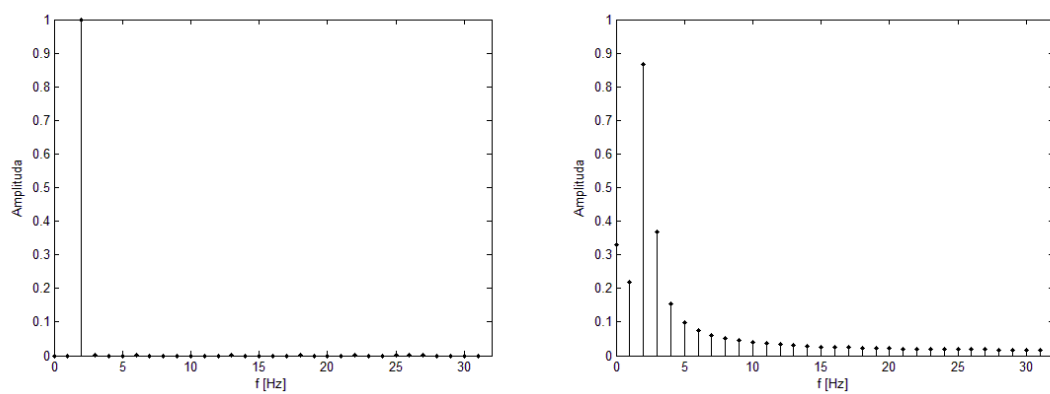
Aplikace okénkovacích funkcí je postup, kterým se snažíme zmírnit tzv. rozmazání spektra. Okénkovací funkci lze aplikovat buď v časové oblasti prostřednictvím pronásobení prvků dvou vektorů, nebo ve frekvenční oblasti operací konvoluce. Jak bylo naznačeno dříve, k rozmazání spektra dochází vždy, když zkoumaný vektor nezachycuje přesně celočíselný násobek periody signálu. To vede k posunu polohy frekvence ve spektru mezi diskrétní hodnoty frekvencí. Pro demonstraci byly vytvořeny dva signály. Oba průběhy jsou vzorkovány s $f_s = 64$ Hz. Zkoumaný frame je dlouhý 1 sekundu – 64 hodnot. První harmonická funkce má frekvenci $f_1 = 2$ Hz, druhá $f_2 = 2,3$ Hz. První dva grafy zachycují časové průběhy těchto signálů a jejich „domyšlené“ periodické pokračování. Druhá dvojice grafů zachycuje amplitudová frekvenční spektra bez aplikace okénkování. Třetí dvojice pak zobrazuje amplitudové frekvenční spektrum po aplikaci Hamingova okénka.

Zjednodušeně lze popsat působení okénkovací funkce v časové oblasti následující úvahou. Problém způsobuje špatné navázání skutečné a „domyšlené“ periodické části signálu. Okénka jsou proto navržena tak, aby váha signálu v jeho „krajních částech“ byla snížena a váha uprostřed se naopak zvyšuje. Výsledkem je hladší navázání skutečné a fantomové složky, což vede k redukci rozmazání spektra.

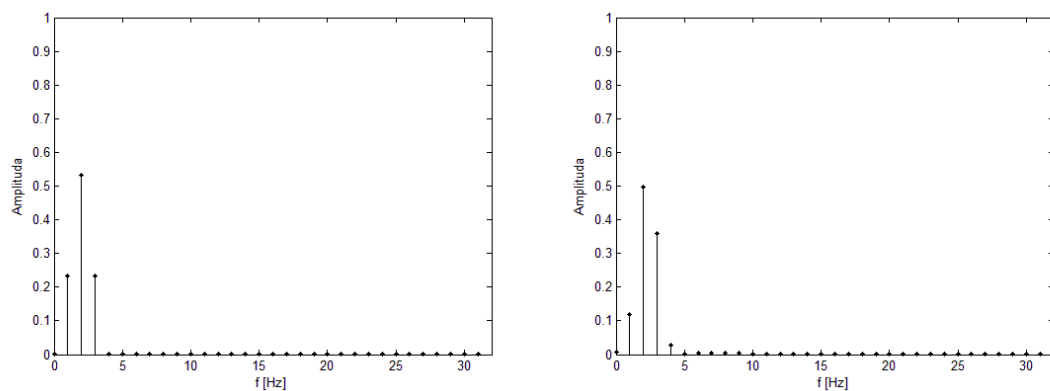
Z grafů (Obr. 2 – Obr. 4) je jasně vidět, že aplikace okének na rozmazané spektrum zmírňuje velikost rozmazání. S okénkováním jsou spojena i určitá negativa. Při aplikaci okénkovací funkce na nerozmazané spektrum dochází k určitému rozmazání. Okénkování zkresluje amplitudy. To se projeví dvěma způsoby – jednak dochází k chybě určení amplitudy až 30%. Velikost této chyby je závislá na typu použité okénkovací funkce. Druhým důsledkem okénkování je změna celkové energie signálu, respektive deformace výkonového spektra.



Obr. 2: Časové průběhy signálů



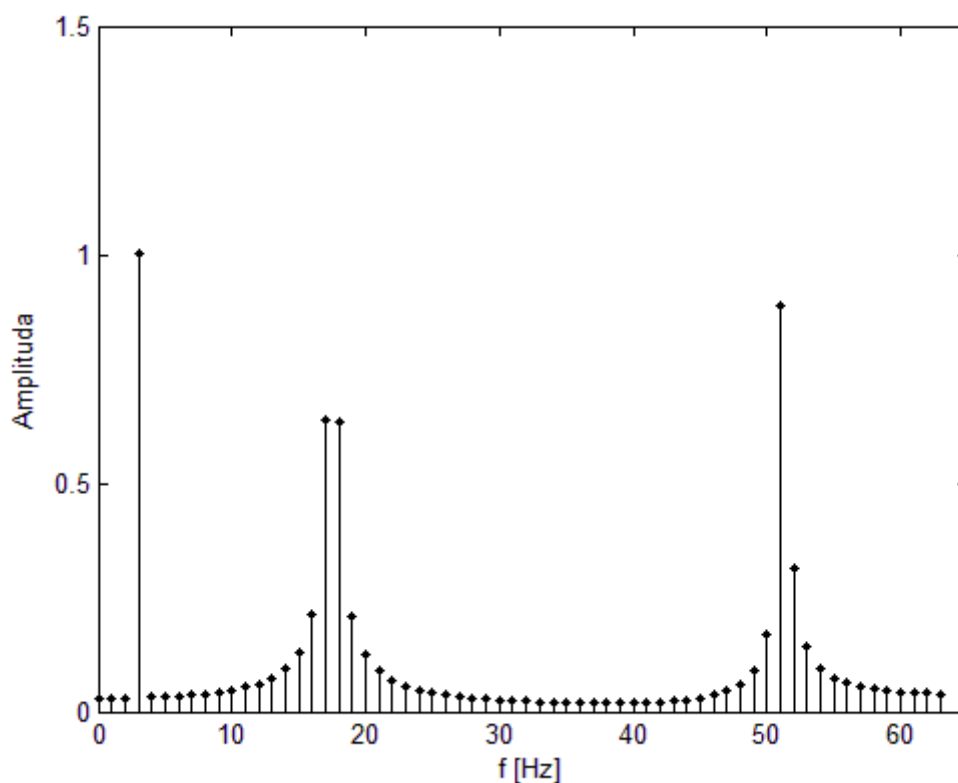
Obr. 3: Amplitudová frekvenční spektra signálů



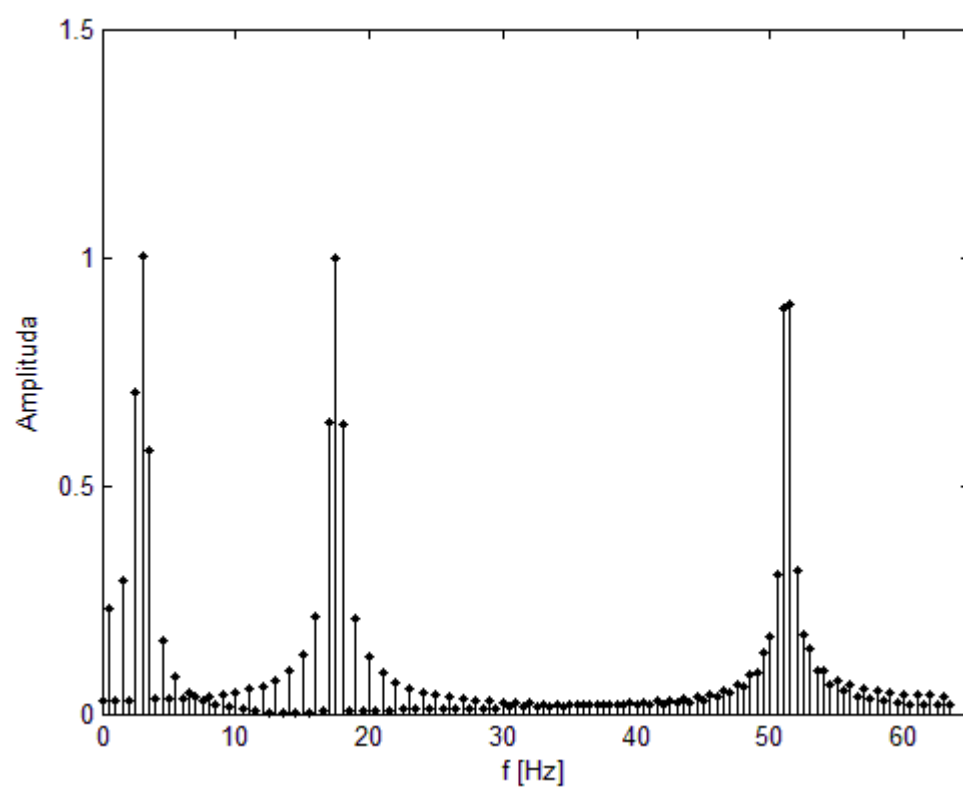
Obr. 4: Amplitudová frekvenční spektra signálů – po okénkování

5.3 Doplnění signálu nulami

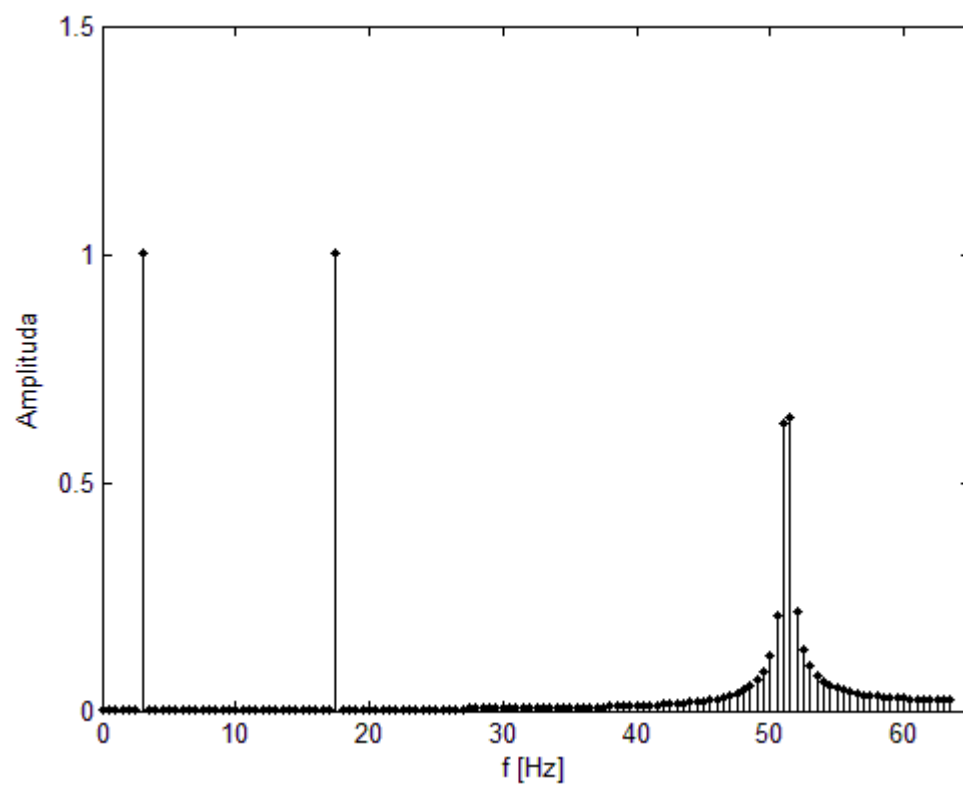
Jak již bylo zmíněno dříve, rozlišení frekvenčního spektra (vzdálenost sousedních spektrálních čar) závisí na délce záznamu a na vzorkovací frekvenci: $\Delta f = f_s / N$. Může se však stát, že takové rozlišení spektra nestačí. Možností, jak zjemnit rozlišení spektra, je doplnit signál nulami. Tím se zvýší celková délka zkoumaného vektoru čísel, zatímco ticho představované nulami by nemělo příliš ovlivnit výsledné amplitudové spektrum. Pro větší názornost byl vygenerován signál obsahující tři harmonické funkce o frekvencích $f_1 = 3$ Hz, $f_2 = 17,5$ Hz a $f_3 = 51,25$ Hz. Signál je vzorkován s $f_s = 128$ Hz a je dlouhý $N = 128$ hodnot. To znamená, že f_2 a f_3 by měly být rozmazané a určení jejich přesné frekvence bude zatíženo chybou. Pokud bychom však zdvojnásobili délku vektoru hodnot N , měla by být přesně určena i f_2 , protože se zobrazí přímo na diskretní frekvenci. Následující grafy (Obr. 5 – Obr. 7) ukazují nejprve amplitudové spektrum neupraveného signálu, pak spektrum signálu doplněného nulami a nakonec spektrum signálu, u kterého bylo vygenerováno $N = 256$ hodnot.



Obr. 5: Amplitudové frekvenční spektrum neupraveného signálu



Obr. 6: Amplitudové frekvenční spektrum signálu doplněného nulami



Obr. 7: Amplitudové frekvenční spektrum signálu dvojnásobné délky

Z grafů je dobře patrné, že signál proložený nulami trpí určitou mírou rozmazání spektra. Pokud na spektrum aplikujeme algoritmus výběru lokálních maxim, je možné určit velmi přesně jak amplitudu, tak frekvenci jednotlivých složek signálu. Lze tedy tvrdit, že doplnění signálu nulami je možné využít ke zvýšení rozlišení spektra při použití framu stejné délky.

5.4 Omezení metod a optimalizace

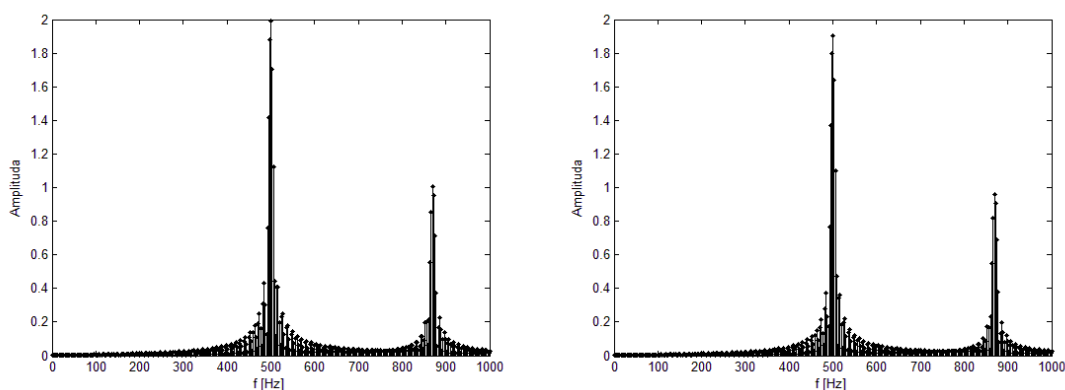
Předchozí kapitoly naznačily určitá omezení použitých analytických metod. Omezení jsou závislá na vzorkovací frekvenci záznamu a na délce zkoumaného framu. Chceme-li analyzovat hudební záznam, je potřeba formulovat požadavky na pořízení záznam, zvolit vhodnou délku analyzovaného framu a případně zapojit další metody, jako doplnění nulami. V závěru kapitoly se zmiňuji o určitých možnostech úspory výpočtů.

Nároky kladené na parametry nahrávky vychází z požadavků na schopnost programu analyzovat určité jevy. Vzhledem k již dříve zmíněným poznatkům z oblasti psychoakustiky, je třeba analyzovat frekvenční spektrum do maximální hrané frekvence 20 kHz. Podle Nyquistova teorému musíme použít vzorkovací frekvenci alespoň dvakrát větší, tedy $f_s > 40$ kHz. Dalším požadavkem je schopnost programu identifikovat šestnáctinovou notu při tempu 200 dob za minutu. Počet vzorků obsazených takovou notou lze určit ze vztahu $N = \frac{60 \cdot f_s}{4 \cdot tempo}$, což pro vzorkovací frekvenci $f_s = 44,1$ kHz představuje $N = 3300$ hodnot, respektive $N = 3600$ hodnot pro $f_s = 48$ kHz. Tuto délku noty lze považovat za mezní případ. Jako vhodnou délku framu tedy použijeme délku $N = 4000$ hodnot. Až do této chvíle jsme ale nebrali v úvahu dolní hranici hledané fundamentální frekvence, která se může pohybovat v hodnotách kolem $f_{min} = 70$ Hz. To by z hlediska Nyquistova teorému nepředstavovalo žádný problém. Je však třeba si uvědomit, že abychom rozeznali dva sousední půltóny, musíme při analýze disponovat dostatečně malým frekvenčním krokem Δf . Z akustické teorie hudby je známo, že podíl frekvencí dvou sousedních půltónů odpovídá vztahu

$$\frac{f_{i+1}}{f_i} = \sqrt[12]{2} \quad (5.5)$$

Je tedy zřejmé, že absolutní velikost tohoto rozdílu je v přímé úměře s velikostí frekvence. Pro frekvence kolem $f_{min} = 70$ Hz je vzdálenost sousedních frekvencí $\Delta f_{min} = 4,1$ Hz. Chceme-li být schopni rozlišit takové frekvence, je nutné dosáhnout v rámci analýzy menšího frekvenčního kroku. V tomto případě to znamená doplnit signál nulami z původní délky framu $N = 4000$ na délku $N = 16384$ hodnot. Konečná délka framu je volena jako mocnina dvojky s ohledem na aplikaci FFT. Pro obě výše zmíněné přípustné vzorkovací frekvence dostáváme frekvenční krok spektra $\Delta f < 3$ Hz.

Navržený algoritmus předpokládá doplnění signálu nulami v poměru cca 3 : 1. Teď se budu zabývat otázkou, zdali je v takové situaci nutné použití okénkovacích funkcí. Doplnění signálu nulami totiž nemusí znamenat pouze „prodloužení“ vektoru. Můžeme přidat polovinu nul před signál a druhou za něj, čímž bude váha okrajových částí signálu nulová, což se nápadně blíží dřívějšímu vysvětlení funkce okének. Na následujících grafech budou zobrazena spektra signálu doplněného nulami, přičemž na první není aplikována okénkovácí funkce a na druhý bylo použito Hamingovo okénko. Z těchto dvou grafů je jasně vidět, že aplikace okénkovacích funkcí je zbytečná, jelikož se na amplitudovém spektru téměř neprojeví.



Obr. 8: Amplitudové frekvenční spektrum signálu neokénkovaného (vlevo) a s aplikací okénka (vpravo)

Podrobnějším rozbořením obou spekter lze zjistit, že aplikace okénkovací funkce mírně zdeformovala amplitudu, zatímco poloha lokálního maxima se nemění. Budeme-li předpokládat použití algoritmů detekce lokálních maxim spektra, je vhodnější okénkovací funkce vůbec neaplikovat.

5.5 Korelace

Korelace je statistická metoda sloužící k určení závislosti dvou znaků souboru $(x_i \text{ a } y_i)$, který má rozsah n . Volíme-li za charakteristiky polohy aritmetické průměry obou znaků $(\bar{x} \text{ a } \bar{y})$ a směrodatné odchylky $(s_x \text{ a } s_y)$ použijeme jako charakteristiku variability, pak míru závislosti znaků popisuje koeficient korelace r_{xy} . Koeficient korelace může nabývat hodnot z intervalu $-1 \div +1$. Hodnota $r_{xy} = -1$ odpovídá nepřímé závislosti znaků, hodnoty blízké $r_{xy} = 0$ značí, že mezi daty není závislost, a $r_{xy} = +1$ značí přímou závislost dat. Korelaci je možné použít jako míru podobnosti dat. Koeficient korelace je dán vztahem

$$r_{xy} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x \cdot s_y} \quad (5.6)$$

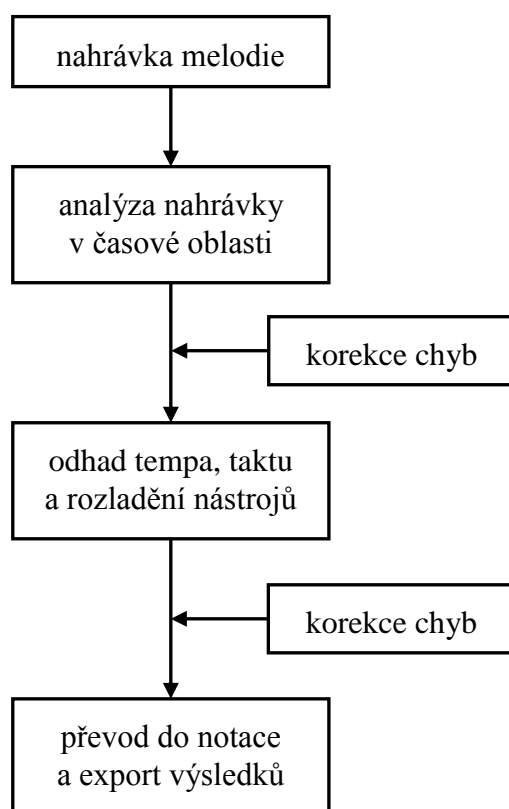
5.6 Energie konečného signálu

Analýza všech typů akustických záznamů s sebou dříve nebo později přináší otázku, jak rozpoznat ticho od signálu nesoucího informaci. Poměrně jednoduchým a robustním nástrojem pro rozlišení pauz a not v hudební nahrávce je energie konečného signálu E . Příkladem aplikace energie signálu, jako příznaku popisujícího signál, je rozdělení záznamu na úseky vhodné délky, které následně porovnáme podle jejich energie. Energie konečného signálu je pro číslicový signál definována jako součet druhých mocnin hodnot vzorků $x[i]$. Díky druhé mocnině hodnot vzorků není velikost energie závislá na smyslu velikosti výchylky, pouze na její absolutní hodnotě. Současně lze říci, že vzorky s vyšší absolutní hodnotou mají vyšší váhu, než vzorky reprezentující nižší výchylku.

$$E = \sum_{i=0}^{N-1} (x[i])^2 \quad (5.7)$$

6 Vrstvy převodu melodie

Převod nahrávky hudební melodie do notace lze rozdělit do několika úrovní, které na sebe logicky navazují. První vrstvou převodu je identifikace melodie v časové oblasti. Po určení časového průběhu skladby navazuje odhad parametrů skladby, jako jsou tempo, takt či míra rozladění nástrojů. Při určení tempa a taktu skladby je vhodnější hovořit spíše o odhadu těchto veličin, protože mohou existovat odlišné zápisy skladby, které vedou na stejný časový průběh (např. dvakrát delší délky not a dvojnásobné tempo). Třetí vrstvou řešení úlohy je samotný zápis časového průběhu do zvoleného formátu notového zápisu. Mezi jednotlivé vrstvy lze vložit algoritmy zaručující korekci některých chyb analýzy (např. opravu nepřesného určení délek not).



Obr. 9: Vývojový diagram zachycující základní uspořádání programu

6.1 Identifikace melodie v časové oblasti

Identifikace melodie v časové oblasti zahrnuje provedení řady dílčích (mnohdy vzájemně nezávislých) analýz a následnou sumarizaci jejich výsledků. Prvním dílčím rozbořem melodie je nalezení intervalů, ve kterých je hrán tón nebo se v melodii nachází pauza. K tomu lze s úspěchem použít rozbor krátkodobé energie signálu. Nalezené intervaly značící přítomnost noty je zapotřebí prověřit algoritmem, který vyhledává začátky dalších not započatých během dozívání not předešlých (tzv. on-set detection). Druhým krokem je určení fundamentálních frekvencí jednotlivých not skladby (ke známé délce noty určíme hraný tón). Známe-li fundamentální frekvence hraných tónů, můžeme přistoupit k určení dalších parametrů not. Prvním parametrem je přítomnost oktávy, druhým pak identifikace nástrojů, které tóny vytvářejí. Součástí této kapitoly je srovnání možností a omezení algoritmů detekce oktáv v závislosti na tom, zda mají k dispozici modely nástrojů.

6.1.1 Rozlišení pauz a not

Náročnost detekce začátků not je velmi závislá na nástroji, který melodii interpretuje. U dechových nástrojů je velmi složité zahrát dvě oddělené noty, aniž by mezi nimi vznikla detekovatelná pauza, která sama o sobě indikuje začátek další noty. Proti tomu se některé skupiny nástrojů chovají výrazně odlišně. Výborným příkladem je klavír. U něj je velmi snadné stisknout opakovaně stejnou klávesu a začít další notu, zatímco předešlá nota dozívá. Dozívání předchozí noty zabrání vzniku detekovatelné pauzy a maskuje fakt, že byla zahrána další nota. Cílem této části práce proto bylo vytvořit natolik robustní algoritmus, aby byl schopen detekovat začátky not pro co největší spektrum hudebních nástrojů. Jako vhodný příznak začátku noty se ukázal výkyv krátkodobé energie signálu. U skutečných nástrojů však dochází k poměrně výraznému kolísání energie. Bylo proto zapotřebí doplnit detekci o nároky kladené jak na strmost výkyvu, tak na jeho velikost vztaženou k průměrným hodnotám dosaženým v okolních segmentech skladby.

Jedním z možných řešení detekce tónů a pauz v melodii je níže popsany algoritmus, který je implementován v programu vzniklém v rámci této diplomové práce.

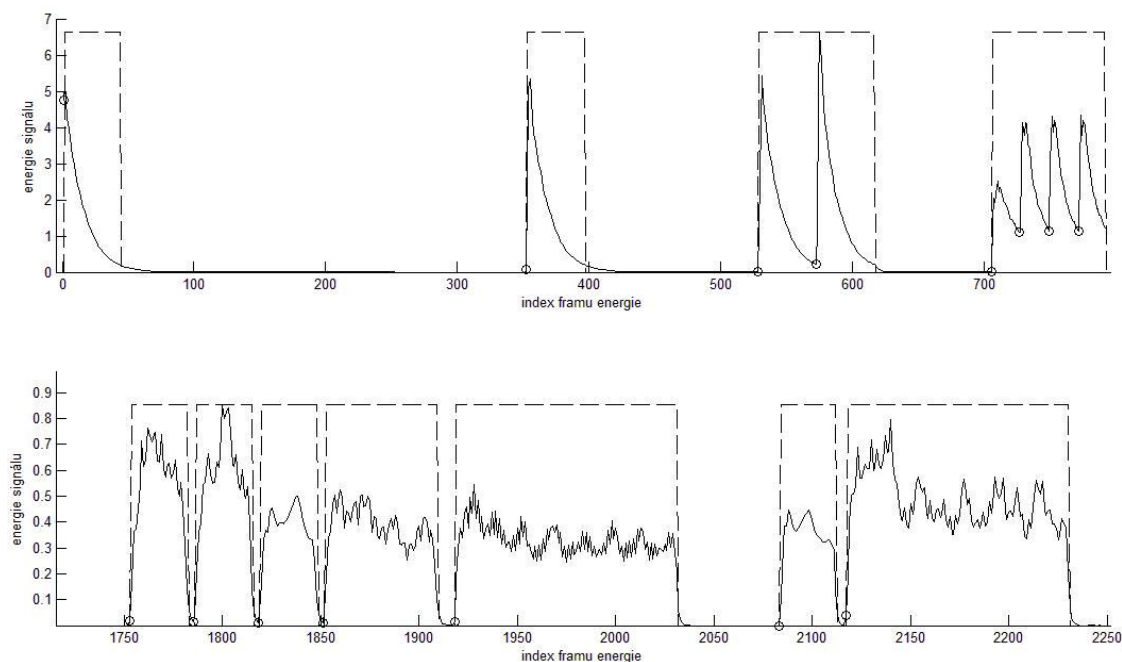
Prvním krokem detekce tónů a pauz je výpočet krátkodobé energie signálu. Energii signálu je potřeba určit na úsecích takové délky, aby bylo zajištěno dostatečné rozlišení intervalů v časové oblasti (co nejkratší segmenty). Současně je vhodné použít dostatečně dlouhé intervaly, aby zajistily určitou míru průměrování energie. Jako vhodné se ukázalo použití segmentů dlouhých 500 vzorků. Z velikosti energie jednotlivých segmentů je určena střední hodnota. Střední hodnota energie je použita jako rozhodovací úroveň pro rozlišení tónu a pauzy tak, že hodnoty energie menší než 30% střední hodnoty energie jsou považovány za pauzu, zatímco hodnoty větší jsou prohlášeny za tón. Hodnota 30% byla určena experimentálně pomocí trénovacích nahrávek skutečné b-trubky a midi-klavíru. Výběr této hodnoty byl formulován na základě zjištění, že již kolem hodnoty 40% jsou výsledky detekce tónů správné a menší hodnota zajišťuje určitou míru robustnosti algoritmu. Z označení jednotlivých segmentů za tón (respektive pauzu) jsou vyhledány bloky souvislých tónů. Nalezené bloky tónů jsou následně kontrolovány na „krátká přerušení“. To znamená, že pokud jeden blok začne dříve než 5 bloků po konci předešlého, předpokládá se, že se jedná o jeden úsek. Pro lepší představu 5 bloků představuje čas 0,057 s. Na takto korigovaných úsecích tónů je následně provedeno dodatečné vyhledání začátků not. Nejprve je na vektor krátkodobých energií aplikován klouzavý průměrovací filtr, který pracuje s daným segmentem a osmi okolními. Tím je dosaženo určité hladkosti průběhu energie nahrávky. Na průměrovaném vektoru energií jsou vyhledána lokální minima a lokální maxima energie. Začátek tónu lze popsat jako výskyt lokálního minima a lokálního maxima v krátkém časovém rozestupu. Aby byl takový jev interpretován jako začátek nové noty, musí splnit ještě dodatečná kritéria zahrnující celkovou velikost skoku a lokální maximum musí dostatečně převýšit předchozí lokální maximum skladby. Výše zmíněné „dodatečné kritérium“ lze formulovat například zápisem:

$$[\max(i) - 2 * \min(i) + \max(i-1)] > A * \max(i) \quad \& \quad (\max(i) - \min(i)) > B * \max(i)$$

Kde $\min(i)$ a $\max(i)$ představují dvojici lokálních extrémů definujících přítomnost nového tónu, $\max(i-1)$ je hodnota předešlého lokálního maxima. Parametr A zastupuje číselnou hodnotu, podle které lze určit, že $\max(i)$ dostatečně převyšuje předchozí signál. Parametr B definuje, zda samotný skok mezi lokálním minimem a následujícím lokálním maximem je dostatečně výrazný. Po řadě experimentů byly použity hodnoty

$A = 0,75$ a $B = 0,3$. Kolem obou hodnot parametrů lze vytyčit toleranční interval $\pm 0,05$. Při volbě parametrů z těchto intervalů se výsledky algoritmu nemění.

Na následujícím obrázku (Obr. 10) jsou vykresleny průběhy energie na segmentu testovací nahrávky midi-klavíru (horní graf) a skutečné b-trubky (spodní graf). Přerušovaná čára znázorňuje bloky nalezených tónů, kolečka představují identifikované začátky not.



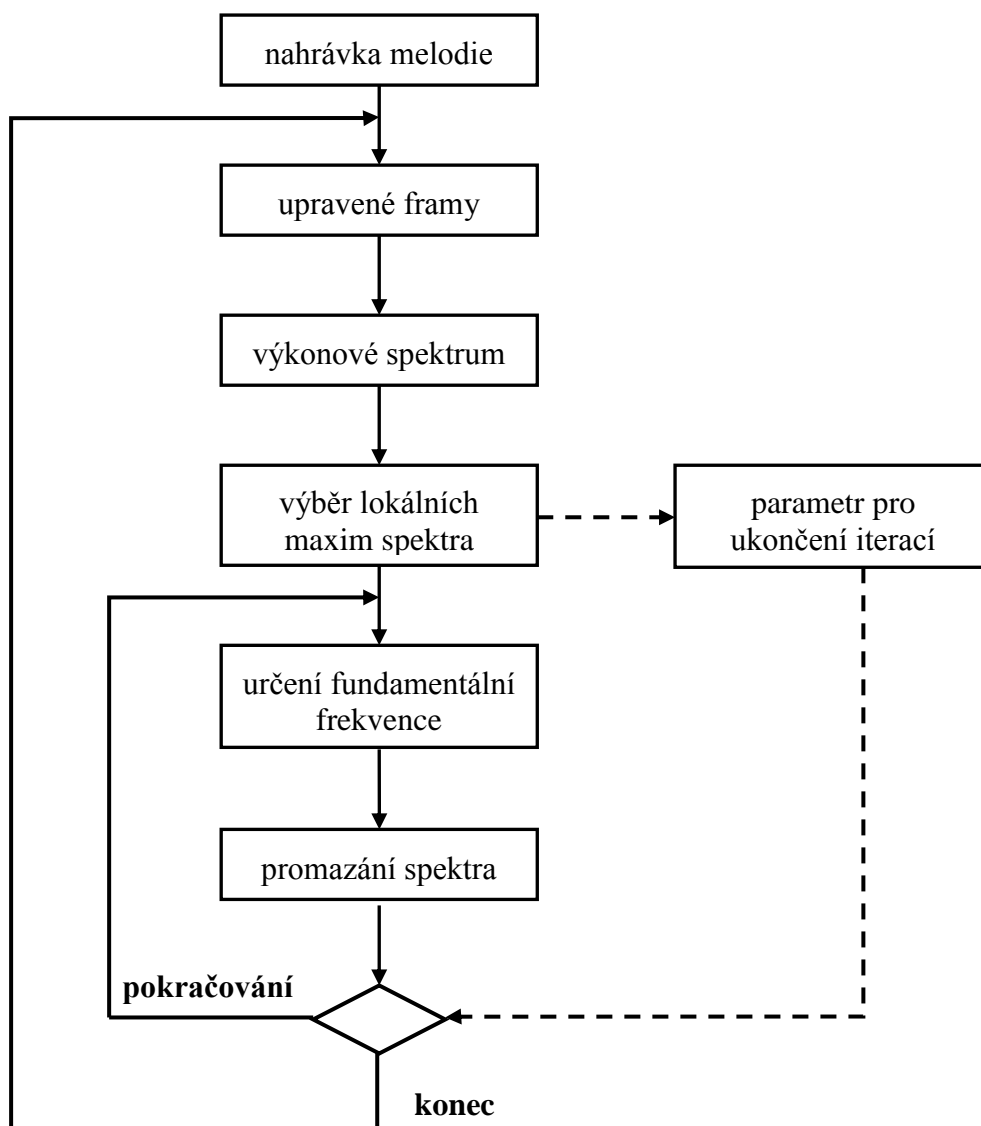
Obr. 10: Průběhy energie signálu midi-klavíru a b-trubky

Z grafů (Obr. 10) lze zjistit, že formulovaný algoritmus je schopen bez chyb nalézt segmenty skladby, kde se nachází pauzy a identifikovat korektně začátky všech not. Za pozornost stojí odlišné průběhy energie signálu po nasazení tónu oběma nástroji. Zatímco energii midi-klavíru (Obr. 10 nahoře) lze popsat jako exponenciálně tlumenou funkci dosahující velkého maxima energie, reálná b-trubka (Obr. 10 dole) má o řád menší maxima energie, ale následující útlum je mnohem méně znatelný. Nezanedbatelné je kolísání energie signálu znějícího tónu.

6.1.2 Určení fundamentální frekvence – PSICE

Základní frekvence generovaná hudebním nástrojem se nazývá fundamentální frekvence. Její hodnota určuje, jakou výšku tónu slyšíme. Všechny ostatní frekvence ve spektru daného nástroje jsou jejími celočíselnými násobky (vyšší harmonické) a tvoří tzv. barvu tónu nástroje. Při analýze vícehlasé hudební melodie musíme předpokládat, že spektrum zkoumaného framu nahrávky obsahuje několik současně znějících různých tónů. V takovém případě jsou ve frekvenčním spektru přítomna spektra jednotlivých tónů, která se vzájemně neovlivňují a lze je díky tomu oddělit. Je zapotřebí identifikovat fundamentální frekvenci každého z nich a přiřadit mu nástroj, kterým byl zahrán. Určitou komplikaci představují situace, kdy by dva současně znějící tóny byly vzájemně posunuty o oktávu. To znamená, že fundamentální frekvence jednoho tónu je dvojnásobkem fundamentální frekvence tónu druhého. V takovém případě se spektra obou tónů překrývají a dochází zde k jejich superpozici. Podobný problém nastává, hrají-li dva nástroje tentýž tón. V rámci řešení této práce byl vyvinut postup, kdy jsou v první fázi identifikovány jednotlivé fundamentální frekvence a detekce oktáv, popřípadě souznějících tónů, je provedena později spolu s identifikací nástrojů.

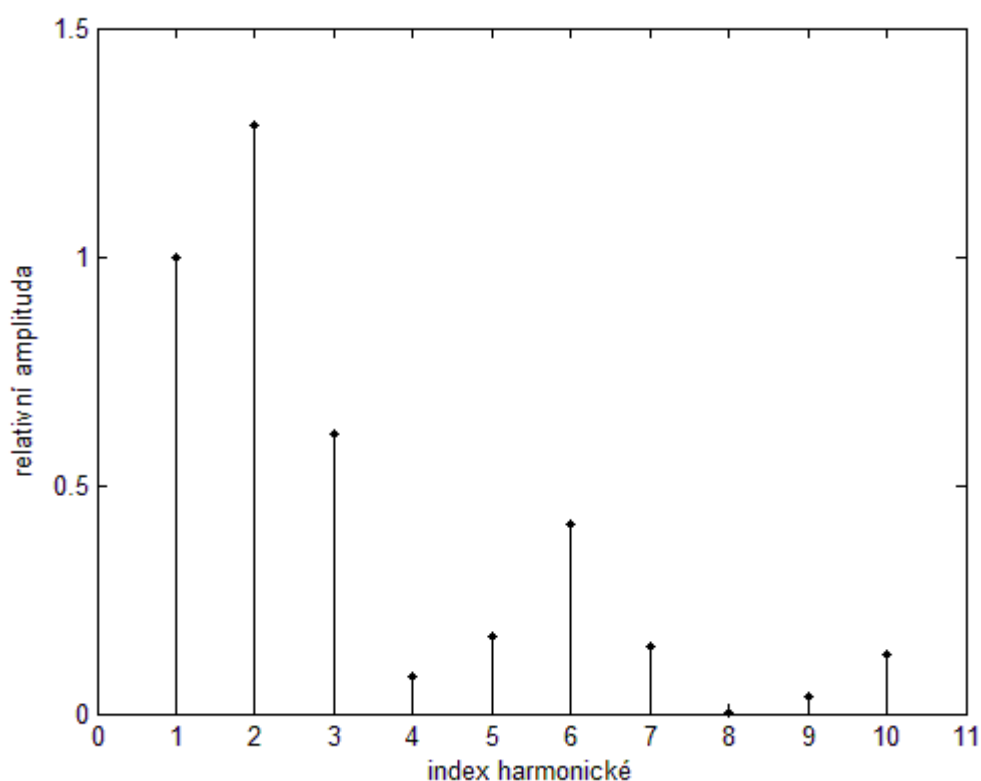
Pro určení fundamentálních frekvencí byl vytvořen algoritmus nazvaný PSICE (Power Spectrum Iterative Cancellation Evaluating). Na první pohled by se mohlo zdát, že PSICE je algoritmus velice podobný postupu „iterative estimation and cancellation“ popsáném v *Signal Processing Methods for the Automatic Transcription of Music* [6]. Hlavní rozdíl spočívá v tom, že A. Klapuri ve svém algoritmu používá model nástroje přímo během analýzy v podobě „koeficientů“, podle kterých odečítá ve frekvenční oblasti části amplitud a části v něm ponechává pro další zpracování. Oproti tomu PSICE je postup, který nepotřebuje znalost modelů nástrojů. Lze ho tedy považovat za robustnější. Následující vývojový diagram zachycuje strukturu algoritmu PSICE tak, jak je implementován v programu. Na jeho začátek je zařazeno i vytvoření framů doplněných nulami. Lze polemizovat, zda algoritmus pracuje přímo s hotovými framy, nebo zda je jejich vytváření jeho součástí. Protože k úpravám framů dochází až po stanovení poloh tónů, rozhodl jsem se tyto úpravy (popsané v 5.4 *Omezení metod a optimalizace*) přiřadit k detekci fundamentální frekvence.



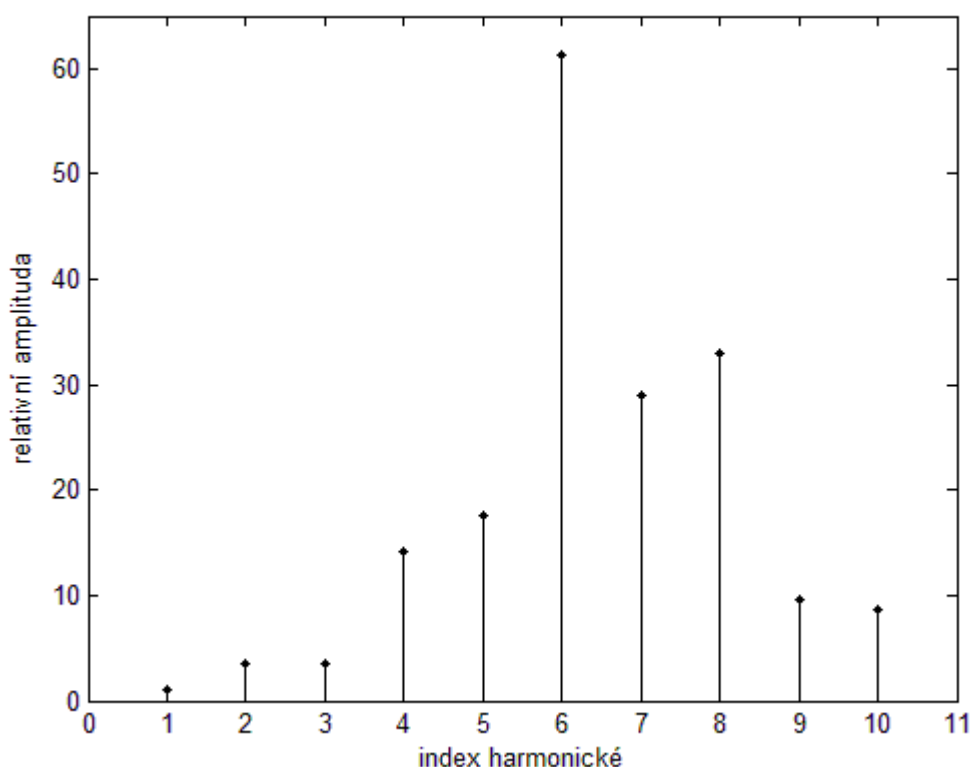
Obr. 11: Vývojový diagram algoritmu PSICE

Určení fundamentálních frekvencí v nahrávce začíná výběrem úseku skladby, který má být analyzován. Každý segment je doplněn nulami způsobem popsáným v 5.4 *Omezení metod a optimalizace*. Na upravený frame je aplikována FFT, z frekvenčního spektra je následně vypočteno výkonové spektrum. Na výkonové spektrum je aplikován algoritmus výběru lokálních maxim. Díky použití lokálních maxim namísto celého spektra se následující postupy stávají méně náročnými na výpočetní čas a navíc je výrazně snížen negativní vliv rozmazání spektra. Podle hodnoty maxima výkonového spektra se určí hodnota lokálního maxima, které je ještě považováno za součást spektra hraného tónu. Tato hodnota rozhoduje o ukončení

iteračního cyklu určování fundamentálních frekvencí. Než přejdeme k popisu samotného určení fundamentální frekvence, je třeba konstatovat několik vlastností výkonových spekter tónů hudebních nástrojů. Nejdůležitější je skutečnost, že fundamentální frekvence nemusí mít ve spektru největší amplitudu. V některých případech je dokonce amplituda výkonového spektra vyšší harmonické řádově větší, než u fundamentální frekvence. Takováto frekvence může být přítomna až na osminásobku fundamentální frekvence. Dobrou ilustrací tohoto faktu jsou amplitudová spektra midi-klavíru a skutečné b-trubky (Obr. 12, 13).



Obr. 12: Normované výkonové spektrum midi-klavíru

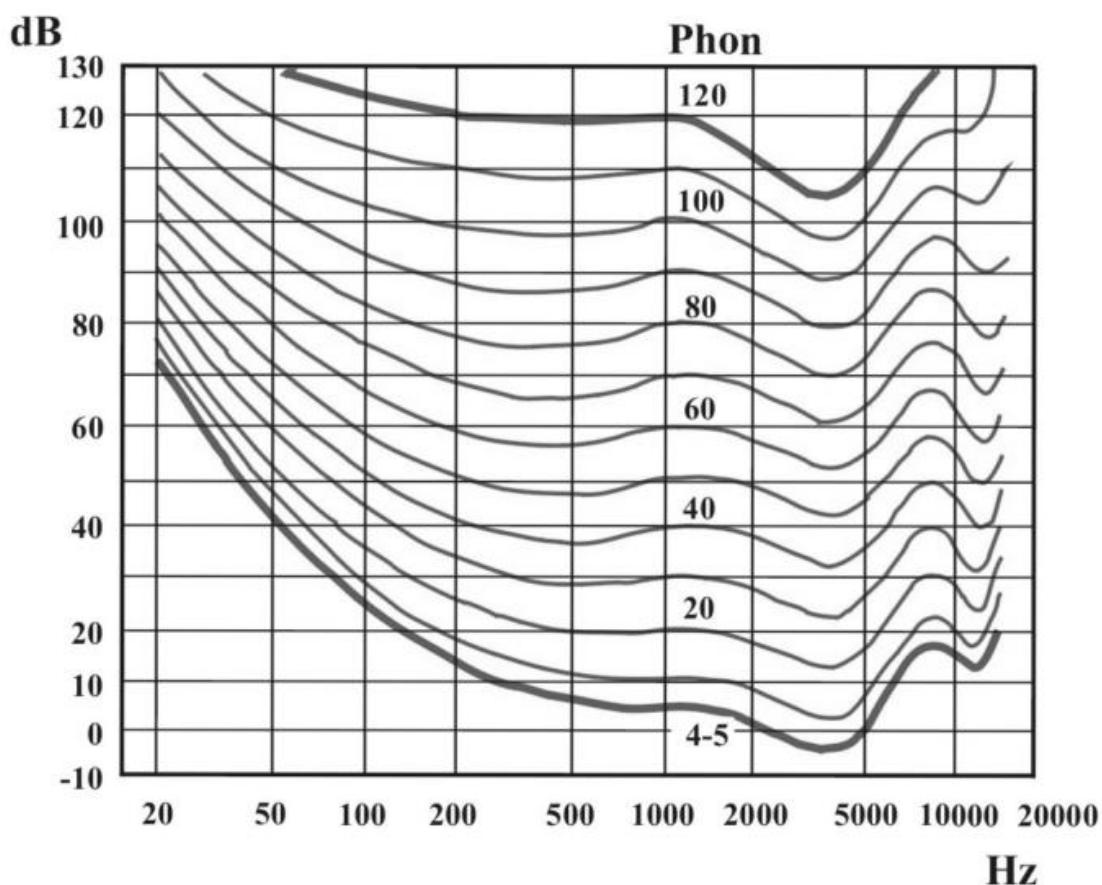


Obr. 13: Normované výkonové spektrum skutečné b-trubky

U některých nástrojů by stačilo prohledávat spektrum od nejmenších frekvencí a ověřit, že k nim existují vyšší harmonické. Tento postup funguje pro nástroje, jako je klavír. Současně má tendenci interpretovat lokální maximum (o amplitudě srovnatelné s šumem) jako fundamentální frekvence. V případě, že frekvence šumové složky je malá, pravděpodobnost nalezení vyšších harmonických ve zbylém spektru je vysoká. Výsledky jsou tak zatíženy značnou chybou. To vedlo k návrhu alternativního postupu, který vychází z předpokladu, že maximum spektra patří buď mezi vyšší harmonické, nebo je samo fundamentální frekvencí. Druhým předpokladem je, že maximum spektra nástroje nebude na frekvenci vyšší, než osminásobek fundamentální frekvence. Prověří se celočíselné podíly frekvencí (respektive úzké toleranční intervaly) na přítomnost lokálního maxima. V případě nalezení lokálního maxima na předpokládaném místě podílové frekvence prohlásí nalezenou složku za fundamentální frekvenci. Po nalezení fundamentální frekvence jsou všechny její vyšší harmonické odstraněny ze spektra a analýza probíhá znovu.

6.1.3 Analýza potřeby modelů nástrojů

Jak bylo nastíněno v předešlém textu, určit fundamentální frekvence vyskytující se v analyzovaném úseku skladby je možné bez apriorní znalosti konkrétních nástrojů. Jediné omezující podmínky deklarované pro algoritmus PSICE byly výskyt maxima frekvenčního spektra nástroje maximálně na osmé harmonické a rozsah fundamentálních frekvencí, na nichž předpokládáme, že se může vyskytnout hraný tón. Identifikovat přítomnost hrané oktávy je možné s použitím znalosti nástrojů – jejich modelů. Studium algoritmu „iterative estimation and cancellation“ navrženým A. Klapurim v *Signal Processing Methods for the Automatic Transcription of Music* [6] může čtenář dojít k závěru, že modely nástrojů nejsou potřeba k detekci oktávy. Aby bylo možné tento rozpor vysvětlit, je třeba uvědomit si několik faktů. Jde hlavně o to, že A. Klapuri analyzuje nahrávky vytvářené jediným nástrojem. Tento nástroj je midi-klavír, pro který je typické, že maximální amplitudu ve výkonovém spektru má buď přímo fundamentální frekvence, nebo druhá harmonická. Když tedy A. Klapuri aplikuje „empiricky určené koeficienty“, které srovnává s podílem amplitud fundamentální frekvence a druhé harmonické, jedná se ve skutečnosti o aplikaci velmi primitivního modelu nástroje. Přesto byl proveden hlubší rozbor metody zkoumání poměrů amplitud fundamentální frekvence a frekvence odpovídající oktávě. K určení rozhodovací úrovně byly využity poznatky z oblasti psychoakustiky, kde je velmi dobře popsána intenzita vnímaného zvuku v závislosti na jeho frekvenci (např. [12]). Taková hranice je frekvenčně značně závislá, což lze doložit na následujícím grafu (Obr. 14). Takto vytvořené kritérium se u midi-klavíru ukázalo jako funkční a dosahovalo velmi vysoké účinnosti detekce oktáv. Problém nastal u jiných nástrojů (jak jinak např. b-trubka). Pro tu se nepodařilo najít ani „empiricky určenou“ sadu koeficientů. Našly se dokonce příklady, kdy u hrané oktávy byla druhá harmonická méně výrazná, než fundamentální frekvence. Možné vysvětlení spočívá ve významu ostatních vyšších harmonických, které se dosud nepopsaným způsobem započítávají do intenzity vnímání nižších frekvencí (včetně fundamentální frekvence a frekvence oktávy). K popsání tohoto jevu by zřejmě bylo nutné detailní pochopení struktury a vzájemných interakcí v rámci sluchového nervu. Dokud tento mechanismus nebude uspokojivě popsán, je nutné k detekci oktáv užívat modely nástrojů.



**Obr. 14: Závislost vnímání úrovně intenzity zvuku na frekvenci
a úrovni hlasitosti (převzato z [12])**

Je zřejmé, že chceme-li jednotlivým melodickým linkám přiřadit nástroje, kterými byly interpretovány, musí být součástí programu modely všech nástrojů, které se mohou v melodii objevit. Otázkou ale zůstává, zda je vhodné snažit se oddělit detekci oktáv od identifikace nástrojů, nebo tyto úlohy řešit současně. V programu, který jsem vytvořil, jsou otázky přítomnosti oktávy, souznějících tónů a identifikace nástrojů řešeny společně. Než se budeme věnovat popisu postupů navržených k rozpoznání nástrojů a odhalení oktáv, je třeba popsat navržený model nástrojů.

6.1.4 Modely nástrojů

Hudební nástroje jsou komplikované systémy, které na základě nejrůznějších fyzikálních principů vytvářejí tóny. V nejjednodušších případech strunných nástrojů se tón vytváří pomocí mechanických vibrací strun zesílených ozvučnou deskou. Dechové nástroje představují mnohem složitější soustavu. Na jejím vstupu je zdroj signálu (strojek hoboje, nátrubek žesťových nástrojů). Nástroj představuje pouze další systém, který vzniklý zvuk upravuje na základě proměnlivé vlastní struktury (systémy otvorů, popřípadě cuků ovlivňujících délku rezonátoru). Barva tónu je ovlivněna i dalšími faktory jako například hlasitost hraného tónu či jeho výška. Vytvořit model nástroje znamená nalézt uspokojivé zjednodušení, jež zachová informace o soustavě a současně nebude příliš obsáhlé a složité.

Východiskem pro formulování modelu nástroje mohou být poznatky z oblasti psychoakustiky. Je známo, že lidské ucho není schopno postřehnout rozdíl mezi dvěma signály, které se liší pouze vzájemným fázovým posunem svých složek. Můžeme tedy předpokládat, že lze vytvořit model nástroje ve frekvenční oblasti a frekvenční spektrum nemusí brát v úvahu fázi jednotlivých spektrálních složek. Jinými slovy pro identifikaci nástroje stačí znalost jeho amplitudového spektra. Druhý předpoklad vychází z teorie analýzy číslicových signálů. Vyjdeme-li z faktu, že tón generovaný hudebním nástrojem je periodický signál, můžeme tvrdit, že všechny složky signálu jsou vyšší harmonické – leží na celočíselných násobcích fundamentální frekvence. Amplitudové spektrum nástroje lze díky tomu zjednodušit na seznam amplitud vyšších harmonických. Studium amplitudových spekter nástrojů bylo zjištěno, že od desáté harmonické jsou amplitudy vyšších harmonických velmi malé. Poslední důležitou skutečností je závislost barvy tónu na jeho výšce. Aby nebyly modely nástrojů příliš rozsáhlé, bylo učiněno další zjednodušení, které spočívá v zahrnutí charakteristických spekter se vzájemným odstupem tercie. Oproti modelu, který by obsahoval každý půltón, je dosaženo redukce rozsahu modelu na jednu čtvrtinu. V konečném výsledku se dá model nástroje interpretovat jako tabulka. Pro ilustraci si lze uvést model midi-klarinetu (Tab. 2).

Tab. 2: Model midi-klarinetu

<i>f_fund</i>	<i>normované amplitudy výkonového spektra</i>									
82	1	0,0019	0,1720	0,0034	0,0662	0,0263	0,1347	0,0236	0,0653	0,0015
97	1	0,0005	1,0964	0,0138	0,1816	0,0255	0,0671	0,0086	0,0024	0,0003
123	1	0,0000	0,5425	0,0002	0,1693	0,0006	0,0163	0,0001	0,0043	0,0000
146	1	0,0000	0,5742	0,0001	0,1676	0,0013	0,0163	0,0000	0,0040	0,0000
184	1	0,0374	4,8061	0,3328	0,6718	0,0000	0,0044	0,0000	0,0014	0,0000
219	1	0,1424	0,7824	0,0473	0,0023	0,0117	0,0000	0,0000	0,0000	0,0000
260	1	0,0787	0,0152	0,0000	0,0005	0,0004	0,0000	0,0000	0,0000	0,0000
328	1	0,4172	0,0189	0,0538	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
390	1	0,4395	0,0136	0,0309	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
491	1	0,4479	0,0125	0,0203	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

První sloupec tabulky představuje indexy spektrálních čar odpovídajících fundamentálním frekvencím (definuje výšku tónu). Indexy jsou určeny při vzorkovací frekvenci $f_s = 44,1$ kHz. Pro jiné vzorkovací frekvence nahrávky je v programu ošetřen přepočít. Následujících deset sloupců odpovídá hodnotám výkonového spektra fundamentální frekvence a prvních devíti vyšších harmonických. Hodnoty výkonového spektra jsou normovány hodnotou výkonového spektra fundamentální frekvence. Proto obsahuje druhý sloupec modelu pouze jedničky a informaci v něm obsaženou lze považovat za redundantní.

Zajímavým jevem, který lze pozorovat u všech zkoumaných hudebních nástrojů je „ztráta barvy tónu“, která se projevuje s rostoucí fundamentální frekvencí. Amplitudy vyšších harmonických postupně klesají s rostoucí fundamentální frekvencí. Pokles amplitud postupuje od nejvyšších frekvencí k nižším. Jelikož se tento jev vyskytuje u všech zkoumaných nástrojů, smazávají se postupně rozdíly mezi jednotlivými nástroji. Jedním z vysvětlení vzniku tohoto jevu je fakt, že vyšší harmonické, jejichž frekvence překročí 20 kHz, jsou vyfiltrovány již na vstupu A/D převodníku, respektive dostanou se mimo oblast slyšitelnosti lidského ucha.

6.1.5 Identifikace oktáv a nástrojů

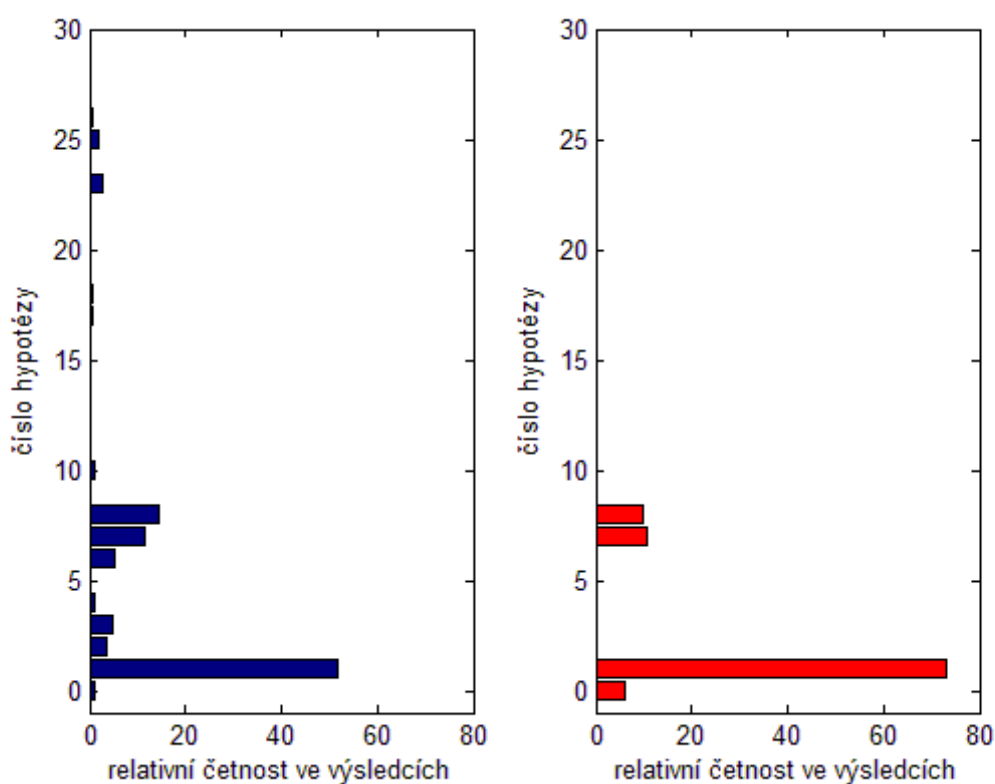
Chceme-li, aby program přiřadil hrané melodii identifikátor nástroje, kterým byla interpretována, musíme mu dát k dispozici modely všech nástrojů, jejichž výskyt lze v nahrávce očekávat. Výzkum v oblasti identifikace hrané oktávy ukázal, že detekce oktáv je možná pouze s použitím modelů nástrojů. Z tohoto důvodu jsou dvě zdánlivě nesouvisející úlohy spojeny v jeden kompaktní algoritmus.

Pro každý zkoumaný úsek skladby je potřeba prověřit řadu hypotéz, jež jsou dány jednotlivými kombinacemi nástrojů, které mohou vygenerovat určené frekvenční spektrum. První skupinou hypotéz je přítomnost jediného tónu hraného jedním jediným nástrojem. Druhá skupina je tvořena předpokladem, že kromě hraného tónu je dalším nástrojem současně hrán tón o oktávu vyšší. Třetí sadou hypotéz, které je zapotřebí prověřit, je možnost, že je dvěma nástroji hrán současně tentýž tón. Hypotézy, kdy různé nástroje hrají různé tóny (mezi tóny není vazba oktávy), jsou separovány již algoritmem PSICE. Další možné hypotézy (například tři nástroje hrající současně tóny s vazbou oktávy) nejsou v programu ošetřeny ze tří důvodů. Jednak je prakticky nemožné takovou kombinaci rozeznat (lidské ucho často selže i při detekci oktávy). Zadruhé jde o fakt, že s takovou kombinací se v reálném hudebním díle téměř nesetkáváme. Třetím důvodem je obrovský nárůst počtu hypotéz, ke kterému by došlo. Souvislost mezi počtem modelů nástrojů a množstvím hypotéz, které generují je rozebrán v kapitole 6.1.6 *Souvislost počtu nástrojů a vzniklých hypotéz*.

Význam jednotlivých hypotéz je hodnocen na základě dvou faktorů. Prvním je korelační koeficient mezi určeným amplitudovým spektrem a amplitudovým spektrem, které by vzniklo použitím nástrojů daných zkoumanou hypotézou. Konkrétní realizace je taková, že od určeného spektra jsou odečteny koeficienty jednoho spektra modelu a zbytek je porovnán s druhým modelem hypotézy. Druhým faktorem užitým při verifikaci hypotéz je „Power Factor“ (**PF**). Ten vypovídá o podílu výkonů mezi spektrem navržené fundamentální frekvence a určeným spektrem zkoumaného framu. Pro případy, kdy je ve spektru přítomna oktáva obvykle nabývá hodnot z intervalu $PF = (0,2; 0,8)$.

Slabina algoritmu spočívá ve skutečnosti, že modely nástrojů neobsahují informaci o fázovém posunu mezi jednotlivými složkami signálu. Místo práce s komplexními čísly jsou použity jejich absolutní hodnoty. Vzniklou chybu lze vymežit pouze určením mezních hodnot součtů pro jednotlivé koeficienty modelu (sčítání fázorů

ve fázi – předpokládáme v algoritmu vs. sčítání fázorů v protifázi – maximální chyba). Tuto chybu lze částečně kompenzovat hromadným vyhodnocením všech framů, které tvoří jeden blok se shodnou fundamentální frekvencí (jeden tón). Cenou za zlepšení výsledků, ke kterému tento statistický přístup vede, je vznik nového omezujícího předpokladu. Pro celý blok hledáme jedinou hypotézu (neměnné složení nástrojů, oktáva nesmí vzniknout během již znějícího tónu). Rozdíl mezi analýzou jednotlivých framů a statistickým vyhodnocením lze ilustrovat následujícími grafy. Na ose x jsou uvedeny procentuální četnosti jednotlivých hypotéz, na ose y jsou indexy jednotlivých hypotéz. Vyhodnocení bylo provedeno na nahrávce midi-barytonu (hypotéza 1).



**Obr. 15: Navržené hypotézy – analýza po framech (vlevo)
a rozbor po blocích framů (vpravo)**

Z grafů (Obr. 15) je jasné patrné, že aplikací analýzy na celý blok zkoumaného tónu lze zcela eliminovat některé chybné hypotézy a snížit podíl ostatních nesprávně určených nástrojů. Celkový podíl správné hypotézy (označena indexem 1) se zvýšil z cca 50% na více než 70%.

6.1.6 Souvislost počtu nástrojů a vzniklých hypotéz

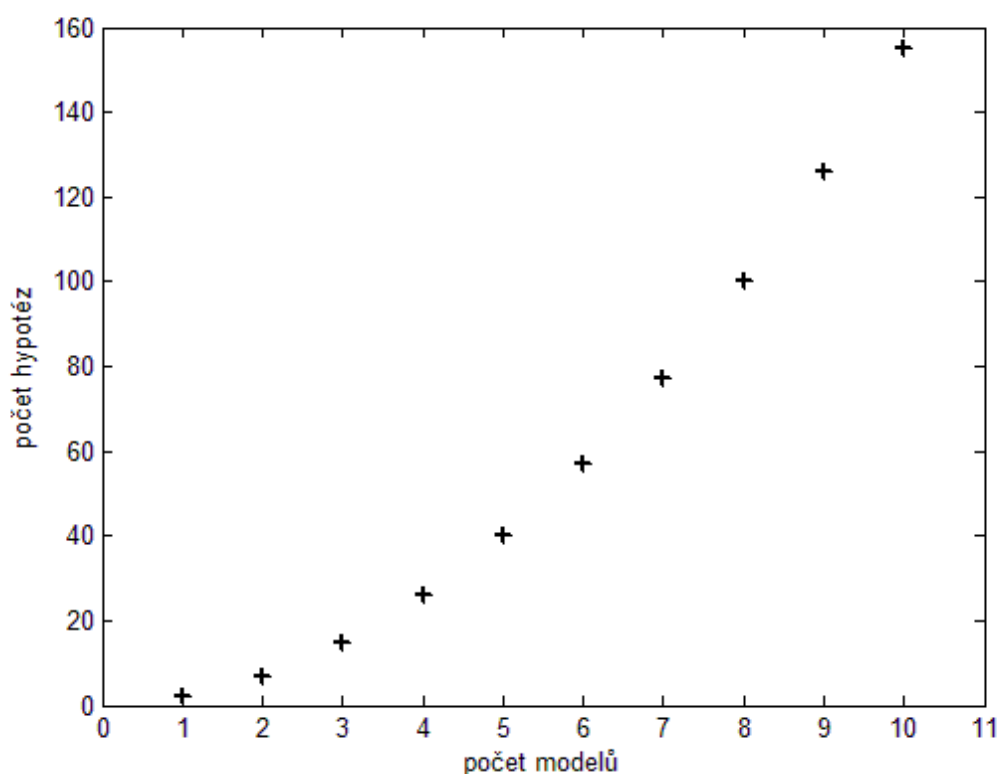
Počet hypotéz o identitě nástrojů a výskytu oktáv je závislý na počtu modelů nástrojů, které má program k dispozici. Počet hypotéz, které je třeba analyzovat, je důležitý ze dvou důvodů. Jednak má počet hypotéz vliv na výpočetní čas potřebný k provedení analýz a jednak lze obecně tvrdit, že čím větší je počet hypotéz, tím větší je i pravděpodobnost chybného vyhodnocení.

Označíme-li N počet modelů a P_i počet hypotéz i -té skupiny, lze celkový počet hypotéz P vyjádřit následujícím způsobem. P_1 zastupuje hypotézy, kdy tón hraje jeden nástroj a v daném úseku se nevyskytuje oktáva. Počet hypotéz v této skupině odpovídá přímo počtu nástrojů: $P_1 = N$. Druhá skupina hypotéz se zabývá hranou oktávou. Počet hypotéz s oktávou je druhou mocninou počtu nástrojů (každý nástroj může znít společně se sebou i s ostatními nástroji), takže $P_2 = N^2$. Třetí skupinu tvoří tentýž tón hraný dvěma nástroji současně. U této skupiny se jedná o výpočet kombinací $P_3 = \binom{N}{2}$. Celkový počet hypotéz $P = P_1 + P_2 + P_3 = N + N^2 + \binom{N}{2}$. (6.1)

Při realizaci programu, který by byl schopen analyzovat „libovolnou“ hudební melodii by bylo zapotřebí použít velmi širokou databázi nástrojů. Například formát MIDI je schopen kódovat 128 různých hudebních nástrojů. Při analýze reálných nástrojů by byl počet modelů ještě větší. V následující tabulce (Tab. 3) jsou uvedeny počty hypotéz pro různé počty modelů nástrojů. Pro počty nástrojů od jedné do deseti je vykreslen graf počtu hypotéz (Obr. 16).

Tab. 3: Počet hypotéz pro různý počet nástrojů

<i>počet modelů</i>	<i>počet hypotéz</i>
5	40
10	155
20	610
50	3775
100	15050
200	60100
300	135150



Obr. 16: Závislost počtu hypotéz na počtu modelů nástrojů

6.1.7 Návrhy rozšíření algoritmů

Během realizace programu byly objeveny možnosti, jak implementované algoritmy rozšířit. Úpravy jsou koncipovány tak, aby zlepšily robustnost algoritmů, popřípadě navrhnou zcela nový postup, který by eliminoval omezení stávajících metod.

První navrhovanou úpravou je rozšíření modelů nástrojů o typický průběh energie signálu, který nástroj vybudí při zahrání tónu. Průběhy energie signálu v čase (během hraní tónu) jsou ovlivněny fyzikálním principem vzniku tónu. U dechových nástrojů má energie signálu průběh velmi podobný obdélníkové funkci. Hlasitost a tudíž i energie signálu závisí výhradně na množství vzduchu, které hudebník vdechne do nástroje. Oproti tomu nástroje, jejichž základem jsou struny (klavír, kytara), vykazují výrazné exponenciální tlumení. Toto rozšíření modelů nástrojů by zpřesnilo výpočet veličiny Power Factor při detekci oktáv.

Druhou úpravou, která by zlepšila výsledky analýzy, je změna detekce on-setů. Stávající algoritmus detekuje začátky tónů podle průběhu energie celkového signálu. Signál je možné rozdělit ve frekvenční oblasti a určovat energii v provázaných frekvenčních pásmech. Provázanost by spočívala v součtu energií v pásmech posunutých vždy o oktávu od sebe (což odpovídá vyšším harmonickým téhož tónu). U stanovení počtu pásem je vhodné použít model lidského ucha. Jako vyhovující počet pásem lze použít 12 nebo 24 pásem na oktávu, kde 12 pásem odpovídá existenci dvanácti půltónů na jednu oktávu, 24 pak bere v úvahu schopnost lidského ucha rozlišit přibližně interval jedné čtvrtiny tónu. Velmi zajímavá metoda určování začátků tónů je prezentována v článku [9]. Společnou vlastností mého návrhu a metody navržené v [9] je řešení detekce v jednotlivých frekvenčních pásmech.

Třetím možným rozšířením je změna struktury modelu a principu rozpoznávání nástrojů. Místo využití modelů, které zachycují amplitudové spektrum nástroje, by bylo možné aplikovat model složený z časových průběhů jedné periody signálu. Periodu je zapotřebí určit pro každou výšku tónu. Výhoda takového modelu spočívá v tom, že časový průběh signálu zachovává informaci o fázovém posunu jeho složek. Obdobný model jsem již s úspěchem použil ve svém semestrálním projektu [2]. Při detekci oktáv by se postupovalo podobně jako u stávajícího modelu. Prvním krokem by bylo určení fázového posunu mezi fundamentálními frekvencemi obou složek signálu. To lze provést na základě předpokladu, že výsledný signál je ve frekvenční oblasti lineární kombinací signálů hypotéz. Pro příslušné složky by bylo možné sestavit rovnice, jejichž řešením by byl jednak fázový posun mezi oběma signály hypotéz také zesílení, se kterým jsou skládány. Další krok by byl stejný jako ve stávajícím postupu – určení korelačních koeficientů mezi zkoumaným úsekem a stanoveným průběhem hypotézy. Power Factor by byl určen přímo z koeficientů zastupujících zesílení jednotlivých složek lineární kombinace spekter (jedná se o aplikaci principu superpozice). Tento model má současně několik nevýhod. Jednak obsahuje mnohem více dat než model s amplitudovými spektry. Druhou nevýhodou je, že skutečné nástroje nebudou mít stejné fázové posuny složek, i když mají stejné výkonové spektrum. Vedlo by to k rozšíření počtu modelů nástrojů – každý nástroj by měl v databázi několik alternativních modelů, což by radikálně zvýšilo nároky na výpočetní čas.

6.2 Odhad tempa a taktu

Máme-li k dispozici časový průběh melodie, je pro převod do hudební notace nutná znalost tempa skladby a taktu, v jakém má být zápis proveden. Přírodním zdrojem informace o tempu skladby je frekvence úderů bicích nástrojů. Odlišením různých nástrojů ze sestavy bicích lze získat informaci o taktu (první tzv. těžká doba bývá hrána na jiný nástroj než zbylé údery taktu). V kontextu této práce předpokládáme malý dechový soubor, u něhož se rytmická sekce nenachází (součástí nahrávek nejsou bicí, klavír ani kytara). Tím je celá úloha ztížena.

Algoritmus implementovaný v programu se proto zaměřil na jiné příznaky nahrávky, které obvykle hudebník k analýze nepoužívá. Prvním příznakem je „míra zaokrouhlení“ délek not. Budeme-li se blížit správnému tempu, lze předpokládat, že rozdíl mezi nalezenou délkou noty a délkou předpokládanou pro odhadnuté tempo bude klesat a po překročení správného tempa se začne rozdíl opět zvyšovat. Tento trend lze očekávat i okolo dvojnásobného či polovičního tempa skladby. Pro výběr správné hypotézy je pak zapotřebí najít další parametr nahrávky. Řešení bylo nalezeno ve zdánlivě nesouvisejícím vědním oboru – v kryptografii.

V 9. století n. l. arabský učenec Abū Yūsuf Ya‘qūb ibn Ishāq al-Kindī (801–873 n. l.), který se věnoval práci knihovníka a překladatele (mnohdy zašifrovaných textů), formuloval následující poznatek. Pro každý jazyk lze vysledovat typickou četnost výskytů jednotlivých znaků v textech. Četnost výskytů znaků je shodná pro libovolné texty psané v daném jazyce. Jeho objev se nazývá frekvenční analýza. Fenomén lze přeformulovat i pro analýzu hudebních nahrávek. Zaměníme znaky jazyka za délky not a jazyk za hudební žánr. Lze ověřit, že četnosti výskytů jednotlivých délek not sledují u stejného hudebního žánru velice podobné rozložení.

Shoda mezi rozložením délek not zkoumané nahrávky a vzorovými rozložením známých žánrů je parametrem, který umožnil vybrat správnou hypotézu tempa (řeší problém dvojnásobného, respektive polovičního tempa). Shrnutí frekvenční analýzy dvaceti trénovacích a deseti testovacích nahrávek je uvedeno v *příloze 1: Frekvenční analýza délek not*. Pro celou sadu skladeb byly odvozeny čtyři žánry. Koeficient korelace mezi skladbou a některým žánrem pak má u všech nahrávek hodnotu 0,9 a vyšší. Maximální korelaci vykazují žánry odpovídající žánru nahrávky.

I skladby, jejichž žánr není přímo uveden mezi žánry vzorovými, mají mezi vzory svůj protějšek, pro který dosahují vysoké míry korelace.

Použití frekvenční analýzy má však mnohem větší potenciál. Samotná shoda četností jednotlivých délek not nestačí na určení žánru. Informaci o typických četnostech not jednotlivé délky lze rozšířit o obvyklý rozsah temp, kterých skladba daného žánru nabývá. Popřípadě je možné sledovat i typické rytmické útvary. Takovýto nepříliš rozsáhlý soubor dat by měl postačit k rozlišení různých tanců či odlišení hudebních stylů.

Druhou informací nutnou k převedení časového průběhu melodie do notového zápisu je takt skladby. Takt se udává ve tvaru zlomku, kdy jmenovatel značí délku noty připadající na jednu dobu. Čítatel pak udává, kolik dob připadá na jeden takt. Pokud úlohu určení tempa označíme za nejednoznačně řešitelnou, určení taktu má téměř neomezený počet řešení. Jediným vodítkem může být fakt, že zřídka kdy přesahuje tón z jednoho taktu do druhého. Možným řešením určení taktu je určení počtu přesahů not mezi takty a volba taktu s nejmenším počtem přesahů. V případě shody přesahů u více hypotéz lze použít statistickou znalost výskytu jednotlivých taktů.

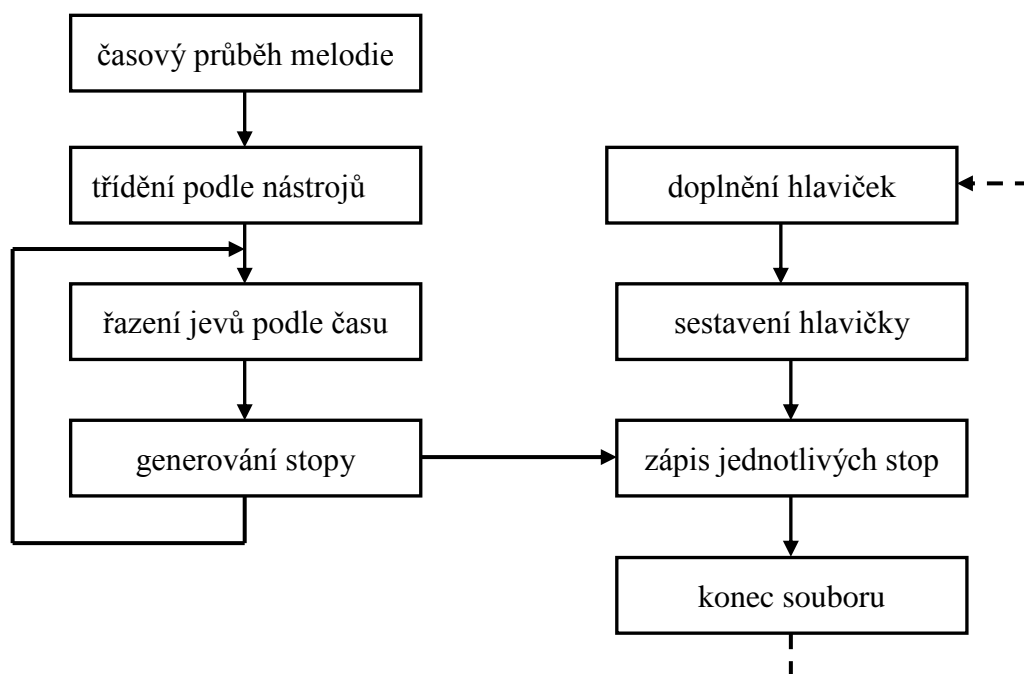
Tab. 4: Porovnání skutečného tempa a taktu s navrženými hodnotami

<i>skutečné tempo</i>	<i>skutečný takt</i>	<i>navržené tempo</i>	<i>navržený takt</i>
110 bpm	4 / 4	120 bpm	4 / 4
120 bpm	4 / 4	120 bpm	4 / 4
100 bpm	4 / 4	110 bpm	4 / 4
100 bpm	4 / 4	100 bpm	4 / 4
120 bpm	4 / 4	120 bpm	4 / 4
90 bpm	3 / 4	90 bpm	4 / 4
100 bpm	4 / 4	110 bpm	4 / 4
100 bpm	4 / 4	100 bpm	4 / 4
100 bpm	3 / 4	100 bpm	4 / 4
80 bpm	4 / 4	75 bpm	4 / 4

6.3 Převod melodie do notového zápisu

V okamžiku, kdy máme k dispozici časový průběh melodie, známe tempo a takt skladby a určili jsme nástroje, které melodii interpretují, je možné přistoupit k převodu získaného popisu skladby do notového zápisu. Protože formát MIDI umožňuje v podstatě jakékoli délky not, zatímco hudební notace se omezuje na určité násobky a podíly délky doby, je třeba realizovat určité korekce určených délek not. V tomto bodě výpočtu je vhodné aplikovat nejrůznější metody korekce chyb. Příkladem korekce chyb je algoritmus hledání opakujících se částí melodie (lajtmotivu) a oprav přeslechů navržený v článku [9].

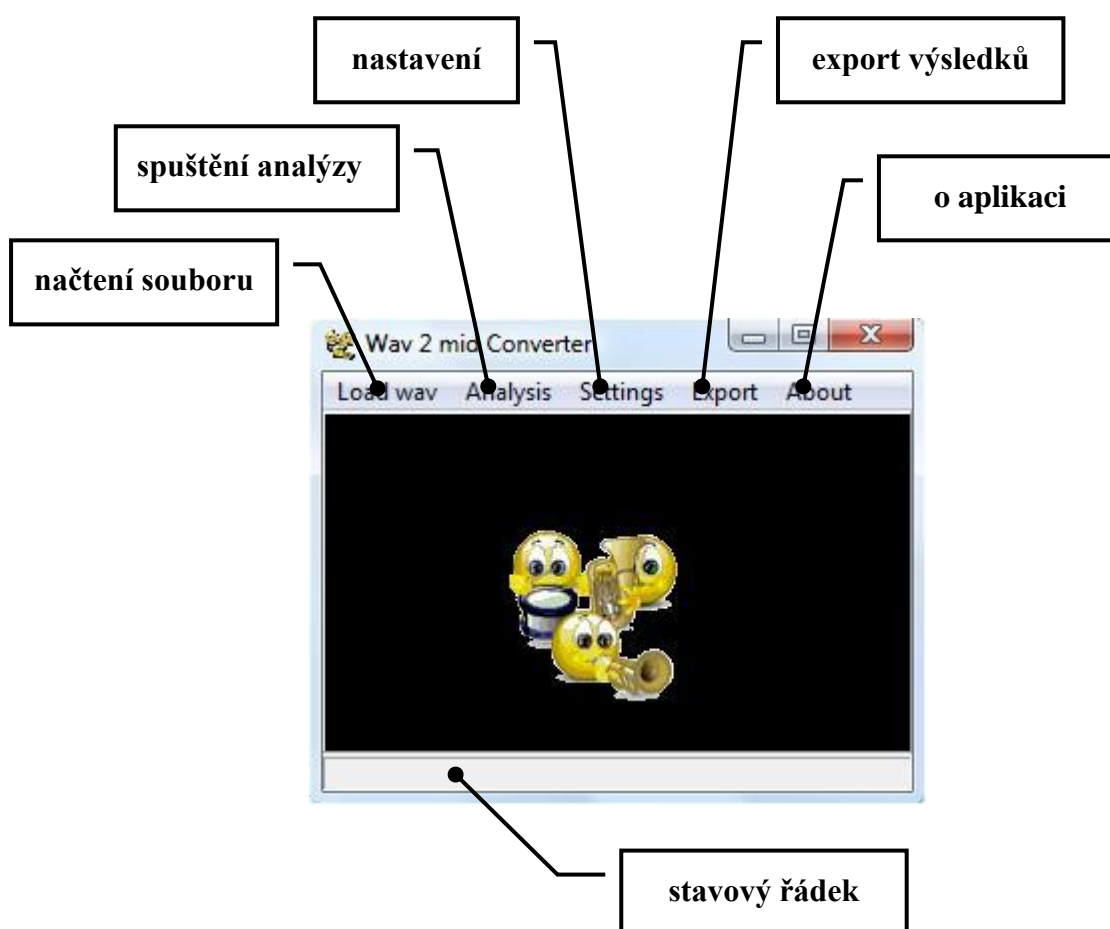
Samotný postup generování souboru MIDI použitý v programu pak spočívá v následujících krocích. Nejprve jsou podle nástrojů separovány jednotlivé party. Každý part je postupně zapsán jako jedna stopa souboru. Korekce délek not je provedena pomocí přiřazení noty do tolerančního intervalu kolem ideální délky noty. Konkrétní implementaci zachycuje vývojový diagram (Obr. 17).



Obr. 17: Generování souboru MIDI

7 Manuál k programu

Program vytvořený v rámci této práce je koncipován tak, aby umožnil analýzu hudební nahrávky s minimálními požadavky směrem k uživateli. Celé ovládání je zprostředkováno nabídkami menu v horní části okna. V manuálu jsou postupně rozebírány jednotlivé nabídky a význam možných voleb.



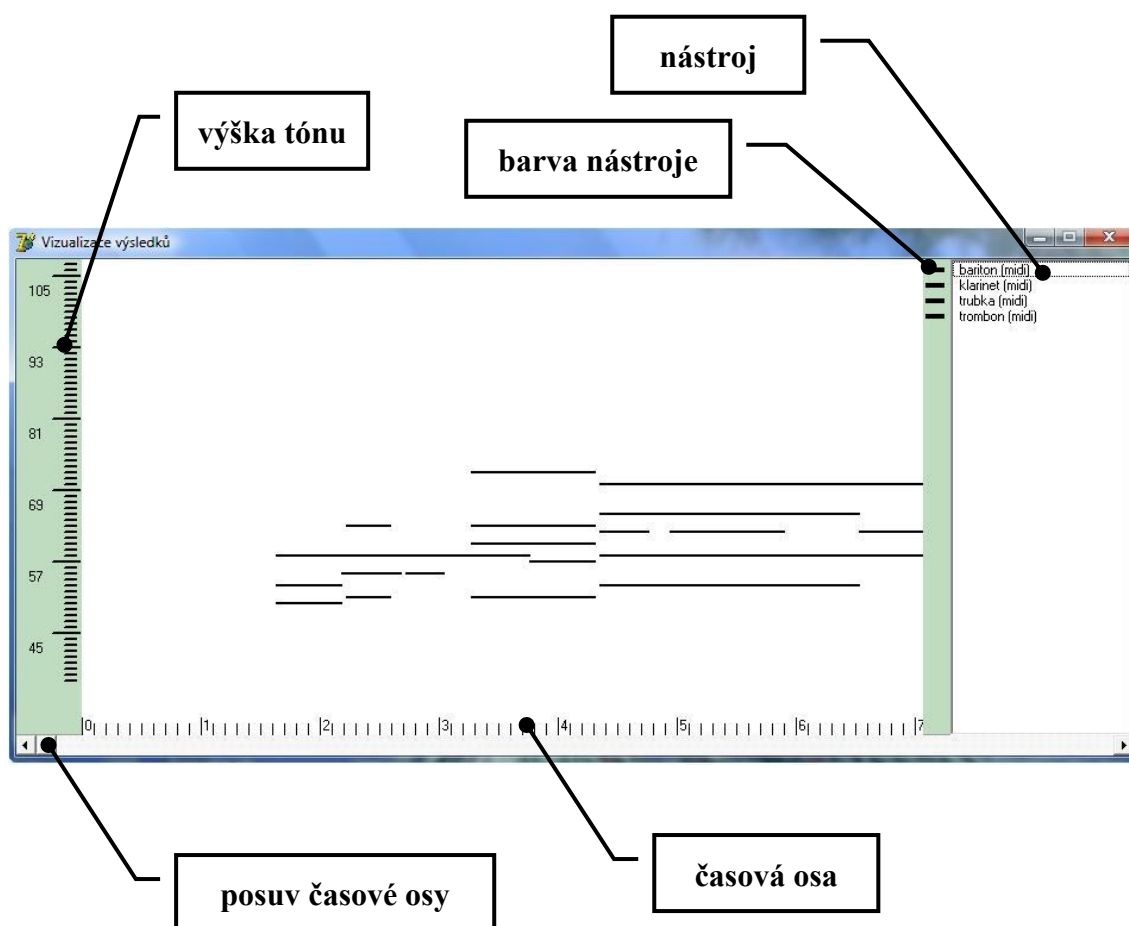
Obr. 18: Základní okno programu

7.1 Načtení souboru

Načtení souboru provedeme pomocí nabídky **Load wav**. Po stisku tlačítka se zobrazí standardní dialog pro výběr souboru. Po načtení souboru se v okně programu zobrazí základní informace o nahrávce.

7.2 Provedení analýzy

Provedení analýzy se aktivuje tlačítkem *Analysis*. Během provádění výpočtu není možné zadávání dalších pokynů. Po dokončení rozboru nahrávky je zobrazen dialog s výsledky (Obr 19). Dialog s výsledkem obsahuje graf, který má na ose *x* vynesena čas v sekundách, na ose *y* je výška tónu (MIDI kód). V pravé části dialogu je seznam nástrojů detekovaných v nahrávce. Kliknutím na obdélník vlevo od názvu nástroje je možné určit barvu, kterou je daný nástroj zobrazen v grafu. Posuvník na spodní straně dialogu slouží k posouvání zobrazeného výřezu časové osy.



Obr. 19: Dialog s výsledky analýzy

7.3 Nastavení parametrů analýzy

Nastavení parametrů analýzy se provádí prostřednictvím záložky *Settings*, kde jsou dostupné volby čtyř parametrů. První položka **model** ukazuje, jaký soubor modelů je právě používán. Soubor modelů obsahuje informaci o modelech nástrojů ve formě normovaného amplitudového spektra a identifikátory nástrojů (slovní název a MIDI kód). Soubor modelu obsahuje i histogramy četností délek not, které jsou použity jako hypotézy při určování tempa (viz. 6.2 *Odhad tempa a taktu*). Pro změnu použitých modelů stačí na položku kliknout a vybrat zde jiný soubor s odpovídající strukturou. Příklad použitého souboru modelů je v *Příloze 2: Použitý soubor modelů*.

Druhou položkou seznamu je **korekce rozladění**. Při práci s reálnými hudebními nástroji se stává, že nástroj je ovlivněn okolním prostředím (např. teplota). V důsledku vlivů okolí se pak frekvence tónů mírně posouvají. Je-li nabídka **korekce rozladění** zaškrtnuta, program vypočítá v rámci analýzy průměrné rozladění nahrávky a provede patřičné korekce. V případě analýzy syntetizované nahrávky není tato korekce potřeba.

Třetí položkou v záložce *Settings* je **kontrola NFV**. Zkratka **NFV** je odvozena z anglického Neighbor Frame Verification. Je-li položka aktivní, je součástí analýzy kontrola, zda se daný tón vyskytuje i v sousedních framech. Nemá-li patřičný tón v sousedních framech nalezen, je výskyt tónu považován za šum a tón je z výsledků vyřazen. Na první pohled by bylo možné toto ověřování používat vždy. Pokud je však nabídka aktivní, znemožní detekci not kratších než šestnáctinová (pro $x / 4$ takt) při tempech vyšších než 70 bpm.

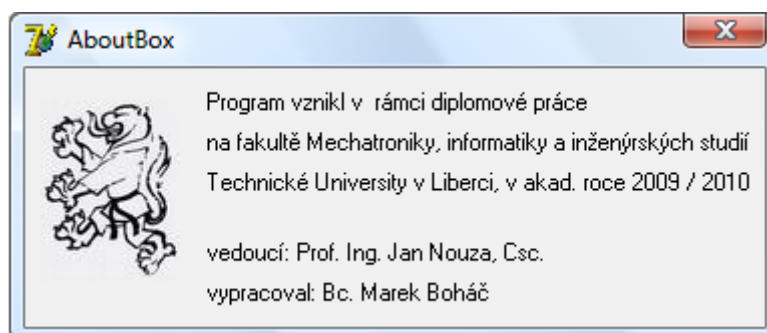
Čtvrtá záložka **identifikace nástrojů** rozhoduje, zda se po algoritmu požaduje, aby identifikoval jednotlivé nástroje a oktávy nebo zda postačuje určení fundamentálních frekvencí v nahrávce a vykreslení bez rozlišení nástrojů. Je-li nabídka neaktivní, je v okně výsledků zobrazen pouze nástroj „NoName“ a při exportu výsledku je stopě přidělen nástroj klavír.

7.4 Export výsledků

Záložka **Export** obsahuje tři položky. První z nich **.mid** umožňuje export výsledků ve standardním formátu MIDI. Druhá položka **.bmp** slouží k uložení grafického výstupu programu ve stejném formátu jako je použit v dialogu vizualizace výsledků. Položkou **tempo / takt** se aktivuje dialog, který umožňuje manuální úpravu tempa a taktu skladby.

7.5 O aplikaci

Záložka **About** aktivuje standardní About box se základními informacemi o programu.



Obr. 20: About Box programu

8 Vyhodnocení úspěšnosti programu

Kapitola je věnována vyhodnocení úspěšnosti programu při převodu hudební nahrávky do notového zápisu. Postupně jsou rozebrány jednotlivé typy chyb, kterých se program při analýze testovacích nahrávek dopustil. Ke každé skupině nepřesností jsou navrženy úpravy algoritmů, kterými je možné nepřesnosti odstranit.

8.1 Odhalené chyby a návrhy řešení

První chybou, které se program dopouští, jsou občasné přeslechy velmi krátkých not. Tyto přeslechy souvisí s kontrolou NVF (popsáno v kapitole 7.3 *Nastavení parametrů analýzy*) a také s přechodnými jevy, ke kterým dochází na počátku tónu. Možným řešením je užití Q-transformace, která (na rozdíl od FFT) zaručuje konstantní počet spektrálních čar na oktávu. Tím se zajistí dostatečné rozlišení mezi nízkými tóny při použití kratší délky framu. Díky většímu počtu framů ve stejně dlouhém úseku skladby se přeslechy krátkých tónů výrazně zredukuje.

Druhou skupinou chyb je nepřesnost určení délek not, která se pohybuje do 10% délky noty. Je způsobena přechodovými ději na začátku tónu (při nasazení tónu dechovým nástrojem dochází ke krátkému ději, jehož fundamentální frekvence neodpovídá hranému tónu) a také zaokrouhlením trvání tónu na framy frekvenční analýzy (rozlišení je omezeno na 0,04 s). Velikost nepřesnosti se pohybuje v rozsahu, který je možné korigovat zaokrouhlením délek při převodu do formátu MIDI. Další možností upřesnění délek not je aplikace pásmové detekce on-setů (popsáno v 6.1.7 *Návrhy rozšíření algoritmů*).

Další skupinu přeslechů lze charakterizovat jako slévání kratších stejně vysokých not, které nejsou odděleny pauzou. Řešení opět spočívá v aplikaci pásmové detekce on-setů. U analýzy skutečných hudebních nahrávek není tato skupina přeslechů tak výrazná, jako u souborů syntetizovaných. MIDI syntetizátor totiž „nasadí tón“ rychleji než skutečný hudebník. U žesťových nástrojů je nasazení tónu podmíněno ucpáním nátrubku špičkou jazyka a následným uvolněním otvoru, což zabírá určitý čas. Mezi po sobě následujícími tóny tak vzniká krátká pauza. Pauza je dost dlouhá,

aby spolehlivě oddělila obě noty. Současně je délka pauzy natolik krátká, že se v notovém zápise neobjeví (časově se pohybuje pod délkou dvaatřicetinové noty).

Algoritmus PSICE nedetekuje některé současně znějící tóny. K této chybě dochází pouze zřídka a je podmíněna nepříznivou kombinací hraných tónů. Když algoritmus nalezne např. fundamentální frekvenci $f_1 = 174,6$ Hz (MIDI tón 53) a současně zní tón o frekvenci $f_2 = 261,6$ Hz (MIDI tón 60), dojde k částečnému promazání spekter. Promazání lze vysvětlit následující rovností $3 * f_1 = 2 * f_2$ (třetí harmonická jednoho tónu leží na stejné pozici jako druhá harmonická druhého tónu). V důsledku částečného promazání spektra pak není f_2 nalezena. Možné řešení spočívá v aplikaci pásmové detekce on-setů a rozšíření vstupu PSICE o seznam „očekávaných frekvencí“, díky kterému by byla f_2 dodatečně nalezena.

Posledním zdrojem přeslechů je chybná identifikace oktáv a nástrojů. Během testování vykazoval algoritmus uspokojivou úspěšnost. Nasazení na skutečnou nahrávku ukázalo velkou závislost na statistickém zpřesnění výsledků, které krátké noty neumožňují. Komplikaci představují přechodné děje na začátku tónu, které vnáší další nepřesnost. Řešení spočívá v kompletním přestavění modelů nástrojů tak, aby braly v úvahu fázi signálu. Toho lze dosáhnout buď analýzou časových průběhů signálu, nebo rozšířením modelů o normovanou fázi. Nedostatky modelů obsahujících pouze amplitudové spektrum ukázaly, že vycházet z vlastností lidského ucha při formulaci modelu není vhodné.

8.2 Porovnání úseku předlohy a automatického přepisu

V této části kapitoly je podrobně porovnán úsek analyzované nahrávky s výsledky automatického přepisu. Jednotlivé odchylky jsou označeny a následně okomentovány.



Obr. 21: Referenční nahrávka (nahore) a výsledek automatického přepisu (dole)

První skupina chyb (**I**) obsahuje dva přeslechy. Nota C1 je detekována i po svém skončení v referenční nahrávce. Důvodem je přechodový děj ostatních dvou tónů, který v součtu spekter vytvoří maximum, jehož fundamentální frekvence připadá právě na C1 (vyšší harmonickou spodního tónu A0). Druhým přeslechem je slévání třech po sobě jdoucích čtvrt'ových not stejné výšky do noty půlové s tečkou.

Druhá chyba (**II**), kterou je výskyt osminové pauzy, vypadá na první pohled jako špatné určení délek not. Ve skutečnosti je způsobena „sléváním“ not. Čtvrt'ová nota z předešlého taktu je spojena se čtvrt'ovou notou s tečkou a v důsledku správného zaokrouhlení vzniká osminová pauza (půlové noty prodloužené o půl doby se prakticky nevyskytují).

Třetí chyba (**III**) je opět důsledkem „slévání“ not stejné výšky. Výskyt osminové noty C1, který je důvodem pro náhradu čtvrt'ových not dvěma osminovými (jedná se o hudebně ekvivalentní zápis) je způsoben přechodovými ději popsány v (**I**).

Čtvrtá chyba (**IV**) je částečně chybějící nota C2, způsobená přeslechem tónu.

Závěr

V rámci diplomové práce byly navrženy speciální algoritmy, které umožňují analýzu vícehlasé hudební nahrávky interpretované různými nástroji. Vyvinuté algoritmy byly implementovány do programu, který umožnil jejich otestování. Program byl následně uzpůsoben k testování nahrávek vytvořených počítačovou syntézou, které napodobují malý dechový orchestr.

Úspěšnost programu byla ověřena pomocí deseti testovacích nahrávek (obsahují od dvou do šesti hlasů). Pro odhalené nedostatky algoritmů byly navrženy úpravy a rozšíření programu. Z časových důvodů nebyly navržené úpravy otestovány.

Hlavní přínosy práce spočívají ve zjištění, že k detekci oktáv a identifikaci nástrojů jsou potřebné modely, které obsahují informaci jak o amplitudě, tak o fázi jednotlivých složek. Druhým přínosem je vytvoření algoritmu pro odhad tempa bez přítomnosti úderů bicích v nahrávce.

Další vývoj programu se může zaměřit několika směry. Prvním je implementace a ověření navržených úprav, druhým pak zpracování širší sady modelů nástrojů. Třetí možnost rozvoje spočívá v zapracování „vědomostní základny hudebníků“ jako např. znalosti akordů, frekvenční analýzu společných výskytů různých nástrojů nebo použití algoritmů detekujících opakující se segmenty melodie, jak je navrženo v článku [9].

Praktický přínos převodu hudební nahrávky do notového zápisu spočívá v možnostech komprimace hudebních souborů bez ztráty kvality. Navíc je možné nahrávku ve formátech jako MIDI upravovat (např. měnit tempo či upravit nástroje, které melodii interpretují). Takové úpravy mohou být užitečné například při výuce tance. Vytvořené algoritmy mohou být aplikovány i v dalších oborech. Například v technické diagnostice (určení zdroje vibrací, při znalosti typických průběhů vibrací v poruchových stavech).

Seznam zkratek a symbolů

OCR	Optical Character Recognition (optické rozpoznávání znaků)
WAVE	Waveform Audio File Format (souborový formát pro uložení zvuku)
RIFF	Resource Interchange File Format (souborový formát pro výměnu zdrojů)
MIDI	Musical Instrument Digital Interface (číslicové rozhraní hudebních nástrojů)
SMF	Standard MIDI File (standardní soubor MIDI)
DFT	Diskrétní Fourierova Transformace
FFT	Fast Fourier Transform (rychlá Fourierova transformace)
A/D	Analogově-digitální
MP3	zkratka pro MPEG – layer3, kompresní formát uložení zvuku
PSICE	Power Spectrum Iterative Cancellation Evaluating – algoritmus pro určení fundamentálních frekvencí záznamu
$X(\omega)$	signál ve frekvenční oblasti
$x(t)$	signál v časové oblasti
ω	úhlová frekvence [rad/s]
t	čas [s]
Δt	vzorkovací perioda
f_s	vzorkovací frekvence
N	počet hodnot číslicového signálu
T	perioda signálu [s]
Δf	frekvenční krok spektra
x_i, y_i	i-té prvky statistických souborů x a y
n	rozsah statistického souboru
\bar{x}, \bar{y}	střední hodnoty statistických souborů x a y
s_x, s_y	směrodatné odchylky statistických souborů x a y
r_{xy}	koeficient korelace
E	energie konečného signálu
$x[i]$	i-tý vzorek číslicového signálu x
PF	Power Factor – parametr detekce oktáv

Seznam použité literatury

- [1] BOHÁČ, M. *Program pro automatický přepis hudební melodie do not.* bakalářská práce. Liberec : TUL, 2008.
- [2] BOHÁČ, M. *Identifikace hudebního nástroje ve zvukové nahrávce.* semestrální projekt. Liberec : TUL, 2009.
- [3] CALDA, E. - DUPAČ, V. *Matematika pro gymnázia – Kombinatorika, pravděpodobnost a statistika.* Praha : Prometheus, 2002.
- [4] FORRÓ, D. *Svět MIDI.* Praha : Grada, 1997. 375 s.
- [5] GRACE, R. *Hudba a zvuk na počítači.* Praha : Grada, 1999. 288 s.
- [6] KLAURI, A. *Signal Processing Methods for the Automatic Transcription of Music* PhD Thesis. Tampere : Tampere University of Technology, 2004.
- [7] NEVŘIVA, P. *Analýza signálů a soustav.* Praha : BEN, 2000. 671 s.
- [8] TŮMA, J. *Zpracování signálů získaných z mechanických systémů užitím FFT.* Praha : Sdělovací technika, 1997. 174 s.
- [9] CONSTANTINI, G. a kol. *Proceedings of MELECON 2010.* Malta, 2010. s. 196 - 201
- [10] Dostupné na www: http://www.aima.cz/jak_slysime.htm [cit. 2010-03-07]
- [11] Dostupné na www: http://en.wikipedia.org/wiki/File:Anatomy_of_the_Human_Ear.svg [cit. 2010-03-07]
- [12] Dostupné na www: http://fu.ff.cuni.cz/vyuka/akustika/3_psychoakustika.pdf [cit. 2010-03-30]

Seznam příloh

Příloha 1: Frekvenční analýza délek not

Příloha 2: Použitý soubor modelů

Příloha 3: CD

Příloha 1: Frekvenční analýza délek not

**Tab. 1: Přehled skladeb s žánrem a korelací
s ostatními žánry (1-20 učební nahrávky, 21-30 testovací)**

číslo	autor / název	žánr	korelace H1	korelace H2	korelace H3	korelace H4
1	V. Vačkář: Vzpomínka na Zbiroh	moderní (H4)	0,572	0,924	0,420	0,720
2	Ch. Gounod: Ave Maria	klasická (H4)	0,722	0,959	0,614	0,806
3	J. Bönnisch: Love song	moderní (H4)	0,478	0,800	0,419	0,983
4	L. Louigi: Třešňové květy	mambo	0,556	0,946	0,442	0,704
5	F. Mañas: Amoroso	moderní (H4)	0,725	0,912	0,616	0,928
6	J. Matys: Suita (věty 1,2 a 4)	klasická (H4)	0,686	0,911	0,608	0,967
7	R. Nováček: Castaldo	pochod (H1)	0,926	0,480	0,870	0,467
8	J. Hančl: Mužně vpřed	pochod (H1)	0,961	0,564	0,901	0,518
9	V. Vačkář: Šohaj	pochod (H1)	0,972	0,404	0,989	0,434
10	J. Labský: Olympia	pochod (H1)	0,986	0,453	0,974	0,474
11	F. Zita: Fanfárový pochod	pochod	0,534	0,949	0,401	0,788
12	F. Kmoch: Jara mládí	pochod (H1)	0,952	0,334	0,976	0,385
13	F. Kmoch: Kolíne, Kolíne	pochod (H1)	0,967	0,686	0,907	0,611
14	E. Štolc: Šly panenky	pochod (H1)	0,996	0,535	0,988	0,539
15	P. Sousa: Pod hvězdnatým praporem	pochod (H1)	0,948	0,531	0,965	0,522
16	K. Vacek: Pro Jarmilku	polka (H2)	0,648	0,973	0,533	0,915
17	V. Vačkář: Nástup borců	pochod (H1)	0,937	0,270	0,979	0,349
18	J. Strauss: Třesky - Plesky	polka (H2)	0,479	0,991	0,355	0,838
19	ABBA: ABBA Gold	moderní (H4)	0,504	0,987	0,386	0,808
20	J. Haydn: Concerto Es dur	klasická (H4)	0,472	0,925	0,374	0,983
21	G. Zamphir: Píseň osamělého pastevce	moderní (H4)	0,255	0,647	0,222	0,903
22	H. L. Hassler: Intrády I	fanfára (H3)	0,933	0,336	0,984	0,395
23	H. L. Hassler: Intrády II	fanfára (H3)	0,856	0,092	0,895	0,211
24	H. L. Hassler: Intrády III	fanfára (H3)	0,974	0,527	0,985	0,520
25	H. L. Hassler: Intrády IV	fanfára (H3)	0,937	0,334	0,985	0,406
26	H. L. Hassler: Intrády V	fanfára (H3)	0,899	0,611	0,823	0,549
27	Korunovační fanfára	fanfára (H3)	0,989	0,566	0,984	0,549
28	Madrigal	fanfára (H3)	0,928	0,404	0,970	0,452
29	Když jsem k vám chodíval	lidová melodie	0,828	0,080	0,911	0,227
30	Měsíček svítí	lidová melodie	0,950	0,377	0,979	0,425

Tab. 2: Přehled četností délek not v rámci nahrávek
(1-20 učební nahrávky, 21-30 testovací), údaje uvedeny v %

<i>skladba \ délka noty</i>	<i>1/16</i>	<i>1/8</i>	<i>3/16</i>	<i>1/4</i>	<i>3/8</i>	<i>1/2</i>	<i>3/4</i>	<i>celá</i>	<i>delší</i>	<i>triola</i>
1	6,86	42,65	4,90	8,82	9,31	16,67	6,86	0,98	0,00	2,94
2	7,29	51,04	0,00	19,79	4,17	10,42	2,08	5,21	0,00	0,00
3	34,39	31,23	0,00	19,37	3,56	5,93	1,98	1,98	0,00	1,58
4	0,00	78,38	0,00	13,51	0,54	1,08	0,00	6,49	0,00	0,00
5	18,93	37,86	0,00	19,42	4,37	16,99	0,00	1,46	0,97	0,00
6	19,85	32,26	0,00	19,85	3,72	8,93	7,20	3,97	4,22	0,00
7	0,00	25,00	0,00	32,14	3,57	31,43	4,29	2,50	0,36	0,71
8	0,00	27,78	0,00	32,22	6,11	26,11	0,00	3,33	2,22	2,22
9	0,00	21,26	0,00	47,70	9,77	14,94	2,30	2,30	0,57	1,15
10	0,00	23,63	0,00	42,86	2,75	21,98	4,95	0,55	3,30	0,00
11	11,27	41,55	0,00	9,15	8,45	11,27	0,00	9,86	2,82	5,63
12	0,00	17,86	0,00	47,62	8,93	16,67	7,14	0,00	1,79	0,00
13	0,00	34,25	0,00	33,56	2,74	20,55	2,05	3,42	3,42	0,00
14	0,00	29,12	0,00	47,25	2,20	17,58	1,10	2,20	0,55	0,00
15	0,00	29,96	0,00	51,82	7,69	5,26	0,40	2,83	2,02	0,00
16	17,16	41,67	0,00	16,67	9,80	9,80	4,90	0,00	0,00	0,00
17	0,00	15,79	0,00	52,63	2,92	16,37	1,17	4,09	2,92	4,09
18	17,45	72,48	0,00	9,40	0,00	0,67	0,00	0,00	0,00	0,00
19	12,48	56,87	4,11	11,22	6,48	1,11	3,00	2,69	1,74	0,32
20	30,75	41,89	1,07	14,26	3,39	6,24	0,18	0,80	0,36	1,07
21	38,13	24,38	0,00	14,38	1,25	4,38	11,25	6,25	0,00	0,00
22	0,00	19,04	0,00	53,98	6,27	10,12	3,61	6,99	0,00	0,00
23	0,00	8,04	0,00	44,05	4,46	29,17	3,87	9,82	0,60	0,00
24	0,00	28,00	0,00	46,74	7,16	10,95	0,21	6,95	0,00	0,00
25	0,00	19,89	0,00	59,67	1,82	11,50	0,91	4,93	1,28	0,00
26	0,00	31,79	0,00	27,99	0,00	29,89	2,72	6,25	1,36	0,00
27	0,00	29,67	0,00	45,05	6,04	13,19	1,65	4,40	0,00	0,00
28	0,00	24,58	0,00	62,57	3,35	5,59	3,91	0,00	0,00	0,00
29	0,00	3,85	0,00	75,00	3,85	13,46	3,85	0,00	0,00	0,00
30	0,00	20,59	0,00	51,47	5,88	13,24	8,82	0,00	0,00	0,00

Příloha 2: Použitý soubor modelů

4 // počet modelů nástrojů v souboru
61 10 baryton (midi) // MIDI kód nástroje / počet spekter / slovní označení nástroje
36 1 1.39114 0.95584 0.15156 0.29399 0.07587 0.07376 0.01106 0.00678 0.00329
46 1 4.38323 3.76334 0.39932 1.65555 0.14442 0.09833 0.02723 0.00428 0.00129
55 1 3.72480 2.98765 0.32932 0.85452 0.15807 0.07935 0.02043 0.00352 0.00495
65 1 4.73010 3.13029 0.55087 0.14036 0.11788 0.07137 0.02790 0.00529 0.00673
82 1 4.57181 0.28781 0.05565 0.07032 0.03408 0.00495 0.00825 0.00001 0.00001
97 1 3.50842 0.34243 0.07038 0.06151 0.02289 0.00717 0.01256 0.00001 0.00001
123 1 0.50665 0.19563 0.03010 0.02130 0.00494 0.00161 0.00036 0.00001 0.00001
146 1 0.67881 0.21929 0.03926 0.01615 0.00537 0.00057 0.00001 0.00001 0.00001
184 1 0.37974 0.04573 0.01197 0.00275 0.00064 0.00001 0.00001 0.00001 0.00001
212 1 0.20249 0.02250 0.01344 0.00189 0.00040 0.00015 0.00001 0.00001 0.00001
72 10 klarinet (midi)
82 1 0.00185 0.17195 0.00343 0.06618 0.02632 0.13471 0.02356 0.06532 0.00150
97 1 0.00047 1.09639 0.01378 0.18155 0.02551 0.06713 0.00859 0.00239 0.00031
123 1 0.00001 0.54247 0.00018 0.16927 0.00055 0.01627 0.00009 0.00431 0.00001
146 1 0.00001 0.57423 0.00011 0.16757 0.00132 0.01625 0.00001 0.00402 0.00001
184 1 0.03743 4.80609 0.33283 0.67185 0.00001 0.00436 0.00001 0.00143 0.00001
219 1 0.14235 0.78236 0.04729 0.00232 0.01169 0.00001 0.00001 0.00001 0.00001
260 1 0.07870 0.01523 0.00001 0.00048 0.00039 0.00001 0.00001 0.00001 0.00001
328 1 0.41720 0.01892 0.05376 0.00001 0.00001 0.00001 0.00001 0.00001 0.00001
390 1 0.43953 0.01358 0.03087 0.00001 0.00001 0.00001 0.00001 0.00001 0.00001
491 1 0.44788 0.01249 0.02027 0.00001 0.00001 0.00001 0.00001 0.00001 0.00001
57 12 trubka (midi)
61 1 9.16908 10.05715 20.03926 29.05304 23.00139 15.16612 6.09488 2.81525 1.04397
73 1 7.31841 7.98176 16.77255 34.44239 18.38014 12.13960 5.07701 2.27257 0.74027
92 1 8.96829 10.23430 21.15485 39.39600 23.99215 12.57216 5.45279 2.38964 0.69225
109 1 1.09831 5.14123 6.46974 2.34661 2.06722 0.79657 0.35949 0.19399 0.08913
130 1 2.05432 3.81095 2.56263 1.08957 0.50643 0.29941 0.15814 0.05201 0.01850
164 1 1.54219 2.52224 1.44147 0.57153 0.25530 0.13493 0.06326 0.01649 0.00001
195 1 18.19687 14.38243 5.01260 3.21792 1.26390 0.44666 0.00001 0.00001 0.00001
245 1 2.62056 5.17616 1.29295 1.00914 0.43666 0.00001 0.00001 0.00001 0.00001
292 1 2.50798 5.39430 1.09600 0.90497 0.00001 0.00001 0.00001 0.00001 0.00001
368 1 2.66738 6.31550 1.15726 0.00001 0.00001 0.00001 0.00001 0.00001 0.00001
437 1 3.07875 6.36578 0.00001 0.00001 0.00001 0.00001 0.00001 0.00001 0.00001
520 1 0.52143 0.00001 0.00001 0.00001 0.00001 0.00001 0.00001 0.00001 0.00001
58 10 trombon (midi)
31 1 2.44120 19.03681 6.17422 4.32319 5.09289 3.34967 1.17845 1.01250 0.39137
37 1 2.55746 16.24770 5.33634 4.36260 4.30211 3.21128 1.09097 1.04762 0.47505
46 1 2.66663 19.68537 6.13047 7.15151 5.23284 3.12823 1.27778 1.31663 0.52187
55 1 2.12119 17.86456 4.41384 4.59636 4.10548 2.65447 0.94586 0.93844 0.41031
65 1 1.98689 16.58778 5.12602 6.90982 5.18887 3.05628 1.13925 1.18374 0.49977
82 1 18.58701 30.13311 10.76959 4.80543 3.96746 0.53118 0.19804 0.52255 0.38201
97 1 4.01139 7.36125 5.96493 2.99930 0.33172 1.00944 0.33581 0.19815 0.07408
123 1 3.79356 8.09336 5.26768 2.85209 0.29172 0.98638 0.28591 0.21729 0.07105
146 1 4.16352 1.69895 0.29445 0.31608 0.14061 0.03853 0.03454 0.00707 0.00379
184 1 0.50493 0.14640 0.08999 0.05182 0.02402 0.01009 0.00326 0.00273 0.00001
4 // počet histogramů délek not
1.13 26.62 0.00 39.70 5.51 18.22 2.34 3.11 2.00
17.30 57.07 0.00 13.03 4.90 5.24 2.45 0.00 0.00
0.00 23.00 0.00 48.58 4.16 15.77 2.41 5.62 0.46
35.51 41.90 0.27 21.82 4.07 10.62 5.15 3.61 1.39

Příloha 3: CD

Notové zápisy a nahrávky

Složka obsahuje nahrávky použité k experimentálnímu otestování programu. Nahrávky jsou uloženy ve formátech WAVE, MIDI a MSCZ (interní formát kompozičního nástroje).

Program

V této složce je uložen spustitelný soubor programu a všechny zdrojové soubory užívané při jeho tvorbě a kompilaci.

Zpráva

Složka obsahuje text diplomové práce ve formátech DOC, DOCX a PDF.